

# Искусственный интеллект

Перед вами книга по искусственному интеллекту от известного технологического евангелиста Романа Душкина, который два с половиной года собирал материал, отбирал информацию, сортировал, фильтровал, подбирал слова и переписывал разделы, готовя для своего читателя великолепный образец литературы, находящейся на стыке популяризации науки и строгого академического изложения фактов.

В данной книге вы найдёте исчерпывающее описание современного состояния технологий искусственного интеллекта и сфер жизни, где их можно применять. Более того, автор сделал упор на гуманитарной составляющей исследований в области искусственного интеллекта, что выгодно отличает эту книгу от множества других изданий по теме. Также здесь развенчиваются многочисленные мифы об искусственном интеллекте и описывается авторское видение будущего.

Издание будет интересно всем, кто хочет быстро погрузиться в горячую тему искусственного интеллекта, получить базовую терминологию и освоить основные методы.



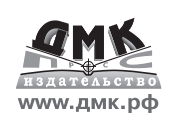
***Роман Душкин*** *- директор по науке и технологиям Агентства Искусственного Интеллекта, член Российской Ассоциации Искусственного Интеллекта. Его курсы по ИИ, машинному обучению, разговорным интерфейсам, философии сознания и другим темам опубликованы на Coursera, Udemy, GeekBrains, а также преподаются в Университете Национальной Технологической Инициативы и Российской Академии Народного Хозяйства и Государственной Службы.*

ISBN 978-5-97060-787-9




**Интернет-магазин:**
www.dmkpress.com

**Оптовая продажа:**
КТК "Галактика"
books@alians-kniga.ru







Искусственный интеллект

Роман Душкин

Искусственный интеллект

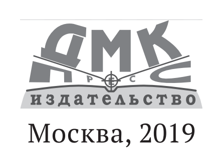


ДМК издательство

AiA АГЕНТСТВО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Р. В. Душкин

# Искусственный интеллект




Москва, 2019

**УДК 004.89**
**ББК 32.813**
**Д86**

Д86 Душкин Р. В.
Исусственный интеллект. – М.: ДМК Пресс, 2019. – 280 с.

**ISBN 978-5-97060-787-9**

Перед вами книга по искусственному интеллекту от известного технологического евангелиста Романа Душкина, директора по науке и технологиям Агентства Искусственного Интеллекта*. В этой книге вы найдёте исчерпывающее описание современного состояния технологий искусственного интеллекта и сфер жизни, где их можно применять. Более того, автор сделал упор на гуманитарной составляющей исследований в области искусственного интеллекта, что выгодно отличает эту книгу от множества других изданий по теме. Также здесь развенчиваются многочисленные мифы об искусственном интеллекте и описывается авторское видение будущего.

Издание будет интересно всем, кто хочет быстро погрузиться в горячую тему искусственного интеллекта, получить базовую терминологию и освоить основные методы.

* Зарегистрированное торговое наименование.

ISBN 978-5-97060-787-9

# ОГЛАВЛЕНИЕ

# ПРЕДИСЛОВИЕ

Как всегда, начиная новую книгу, я испытываю двойственные чувства. С одной стороны, хочется написать по-настоящему полезную книгу, отражающую всё накопленное к текущему моменту знание по заявленной теме. С другой стороны, всегда опасаешься, что полученная в итоге рукопись сразу же окажется устаревшей, ведь сегодня технологии несутся в будущее со скоростью, превышающей возможности отдельного человека. И осознание этого немного удручает.

Тем не менее сегодня я сажусь за новую книгу, в которой постараюсь раскрыть неискушённому читателю тему искусственного интеллекта. Ведь не секрет, что в последние годы вокруг этого понятия поднялась невообразимая шумиха, которая доносится как из средств массовой информации, так и из деловой среды, в которой что ни день открывается новый стартап, основанный на самых современных технологиях искусственного интеллекта. Подогревает эту ситуацию и массовая культура, где тема ИИ тоже стала горячей. Появляется большое количество новых произведений, в которых человеческое общество так или иначе взаимодействует с порождённым (или самозародившимся) искусственным интеллектом.

Впрочем, в киноиндустрии и художественной литературе тема искусственного интеллекта эксплуатировалась с давних времён. Первые писатели-фантасты уже задумывались об искусственных существах, наделённых разумом и пытающихся либо помогать людям, либо столкнуть человека с пьедестала «царя природы» и «венца творения». Дальше – больше. Серия фильмов про терминаторов, затем красочно-философские фильмы «Матрица» необычайно подогрели интерес к этой теме. Художественная литература тоже не отставала. Если не углубляться, то на поверхности лежат трилогия «Киберпространство» Уильяма Гибсона и сага-тетралогия «Гиперион» Дэна Симмонса. Не побоюсь сказать, что только эти два произведения оказали на развитие понимания искусственного интеллекта у массового читателя большее влияние, чем все остальные книги вместе взятые.

Вместе с тем отношение в научной и инженерной среде к термину «искусственный интеллект» и этому направлению вообще

несколько предвзято. Всё дело в так называемых «зимах искусственного интеллекта», т. е. практически полных остановках исследований из-за отсутствия финансирования и разочарования пионеров новой науки. На смену первым восторженным надеждам пришло горькое осознание того, что человек ещё очень далёк от понимания природы сознания и всех тех особенностей мозга, которые делают человека разумным существом. К тому же ещё двадцать лет назад не было достаточных вычислительных мощностей для реализации всех тех теоретических находок, которые были сделаны в научных лабораториях. Так что в среде понимающих специалистов отношение к теме скорее скептическое. Кроме этого, развитие искусственного интеллекта как междисциплинарного направления исследований уже пережило две зимы.

Причиной шумихи вокруг искусственного интеллекта, начавшейся во втором десятилетии XXI века, как видится, являются два процесса. Во-первых, вычислительные мощности и объём имеющихся в распоряжении человечества в целом вычислительных устройств достигли небывалых размеров, и этот размер имеет тенденцию увеличиваться по экспоненциальному закону. Сегодня количество смартфонов, которые можно связать в грид для распределённых вычислений (причём часто без ведома владельцев), достигло двух миллиардов, а каждый смартфон обладает мощностью, на порядки превышающей мощность тех персональных компьютеров, которые были в распоряжении учёных ещё двадцать пять лет назад.

Во-вторых, выросло, получило образование и начало усердно работать поколение людей, которые застали вторую зиму искусственного интеллекта еще младенцами. Но сегодня, получив образование, намного более серьёзное, чем предшественники, представители этого поколения с удесятерёнными силами ухватились за старые надежды, пренебрегая тем скепсисом, который всё ещё имеется у представителей «старой школы». Это и хорошо, и не очень. Хорошо потому, что, не будучи зашоренными, новые специалисты могут «перепрыгнуть» барьер недоверия, выстроенный вокруг искусственного интеллекта. Не очень потому, что многие наступят на те же грабли, что и исследователи первой половины XX века.

Из этих двух тезисов и родилась идея книги. Я хотел бы дать полноценное описание современного состояния технологий искусственного интеллекта, чтобы оно стало основой для дальней-

шего развития новых специалистов, которые уже смогли бы сдать в утиль старые парадигмы и пройти узкой тропинкой к созданию настоящего искусственного интеллекта, так как любая разумная раса в конечном итоге должна задуматься о природе своего разума и стремиться к созданию разума искусственного.

В книге мы обсудим все аспекты искусственного интеллекта. Начнём с истории развития этой области знаний. Поговорим о том, как всё начиналось и какие направления были открыты и потом закрыты по мере развития понимания проблемы. Мы изучим все предпосылки, послужившие формированию новых технологий. В конце первой главы мы изучим типы искусственного интеллекта и получим все необходимые знания, для того чтобы перейти к рассмотрению технологий и их современного состояния.

Во второй главе мы кратко рассмотрим три основополагающих современных метода построения искусственных интеллектуальных систем – символьные вычисления и логический вывод, искусственные нейронные сети и эволюционные алгоритмы. Фактически сегодня это три столпа искусственного интеллекта, на которых базируются прикладные направления исследований.

Далее в третьей главе книги будет представлено более или менее подробное описание технологий искусственного интеллекта и их текущее состояние. Мы начнём с двух направлений, которые развивались в рамках исследований по искусственному интеллекту, и постепенно перейдём к исследованиям, находящимся сегодня на острие науки. В разделе рассмотрим такие темы, как машинное обучение, искусственные нейронные сети, символьные вычисления, методы представления знаний, обработка естественного языка, робототехника и бионика и др.

Четвёртая глава посвящена философии искусственного интеллекта. Мы рассмотрим разные модели сознания, изучим философские подходы к изучению его природы. Это позволит понять, насколько мы ещё далеки от понимания того, что представляет собой разум человека и где он может находиться. Вместе с тем изучение философии искусственного интеллекта позволит очертить те направления движения, куда необходимо двигаться для постижения сути интеллекта и осуществления попыток его разработки в виде искусственной системы.

В пятой главе представлены описания и объяснения всевозможных мифов и опасений человечества относительно искус-

ственного интеллекта и его взаимодействия с людьми и всем человеческим обществом в целом. В массовом сознании сформировалось некоторое количество страхов, которые сегодня могут вылиться в попытки государств или даже надгосударственных структур ограничить или как-то регламентировать исследования в этой области знаний, как это ранее было сделано с генетикой и точными медицинскими технологиями. Несомненно, какое-то регламентирование необходимо, но оно должно быть основано на здравом понимании, а не на страхах и мифах.

Шестая глава посвящена вопросам применения технологий искусственного интеллекта в различных сферах жизни – медицине, образовании, обеспечении безопасности и многих других. В разделе описаны как уже имеющиеся применения технологий, так и варианты будущих применений при развитии методов. Этот раздел будет наиболее интересен тем читателям, которые ищут новые идеи и возможности по изменению существующего порядка вещей.

Седьмая глава описывает перспективы и прогнозы технологий искусственного интеллекта. Что они дадут конкретным людям и человечеству в целом, куда приведут нас как биологический вид, и наконец, какие параллельные исследования получат второе дыхание. Мы рассмотрим футуристические предположения о том, что нам даст искусственный интеллект и когда это может произойти.

Наконец, тех, кто прочитал все шесть разделов до конца, в конце книги ждёт небольшой сюрприз.

В добрый путь…

*Москва, 2017–2019 гг.*
*Душкин Р. В.*

# ПРЕДИСЛОВИЕ ОТ ИЗДАТЕЛЬСТВА

## Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв прямо на нашем сайте **www.dmkpress.com**, зайдя на страницу книги, и оставить комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу **dmkpress@gmail.com**, при этом напишите название книги в теме письма.

Если есть тема, в которой вы квалифицированы, и вы заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу **http://dmkpress.com/authors/publish_book/** или напишите в издательство по адресу **dmkpress@gmail.com**.

## Список опечаток

Хотя мы приняли все возможные меры для того, чтобы удостовериться в качестве наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в тексте или в коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от расстройств и поможете нам улучшить последующие версии этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу **dmkpress@gmail.com**, и мы исправим это в следующих тиражах.

## Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконно выполненной копией любой нашей книги, пожалуйста, сообщите нам адрес копии или веб-сайта, чтобы мы могли применить санкции.

Пожалуйста, свяжитесь с нами по адресу электронной почты **dmkpress@gmail.com** со ссылкой на подозрительные материалы.

Мы высоко ценим любую помощь по защите наших авторов, помогающую нам предоставлять вам качественные материалы.

# Глава 1

# ИСТОРИЯ РАЗВИТИЯ

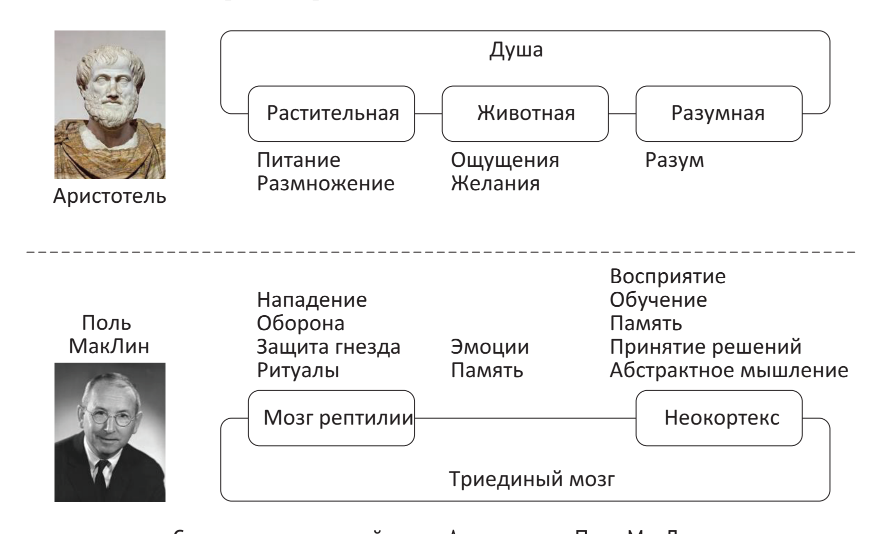
Похоже, что природа разума и сознания взволновала человека прямо сразу, как только он смог осознать свою личность и процесс мышления. Научных свидетельств этому нет, однако можно предположить, что все эти отсылки к «душе», «искре божьей», «духу» и другим подобным штукам имеют в своей основе попытки найти то самое сокровенное, чем человек отличается от «неразумных животных». Конечно же, это относится к разуму, относительно которого до сих пор невозможно сказать как он работает, где скрыт, чем обусловлен. Поскольку подступиться к научному решению этого вопроса до возникновения естественно-научного мировоззрения было невозможно, любопытным людям приходилось плодить сущности.

Вместе с тем развитие научного метода и философии науки происходило в рамках европейской цивилизации, поэтому дальнейшее изложение будет проводиться с европоцентристской точки зрения. Такая позиция не должна отпугивать читателя, поскольку современные достижения науки своей основой лежат именно в европейской колыбели, имея лишь небольшие вкрапления знаний иных цивилизаций. Несмотря на то что такие древние цивилизации, как Индия, Китай и арабский мир, принесли много интересных находок и идей, все они были переосмыслены в европейской традиции, после чего влились в единый поток научного знания.

Поскольку и европейская цивилизация, и современная наука имеют своим истоком древнегреческую философию, имеет смысл обратиться к ней. Особый интерес представляет классический период древнегреческой философии, одним из ярчайших представителей которого является Аристотель. Именно в его трудах складывается научная система мира и познания объективной реальности. Не зря вся современная наука базируется на той самой формальной логике, которую разработал и детально описал Аристотель.

Интересно то, что, согласно Аристотелю, душа человека вмещает в себя три части – растительную, животную и разумную. Первая часть души отвечает за такие функции, как питание и размножение. Вторая часть, животная, – ощущение и желание. Наконец, разумная часть души имеется только у человека, и именно в ней находится разум. Это понимание очень похоже на триединую модель мозга,

предложенную в 60-х годах XX века американским нейрофизиологом Полем МакЛином. В этой модели мозг разделён на три взаимосвязанные и тесно взаимодействующие части – так называемый «мозг рептилии», лимбическую систему и неокортекс. Собственно, мозг рептилии – это самая древняя часть головного мозга, которая имеется у рептилий и птиц и отвечает за базовые функции, связанные с инстинктивным поведением различного плана: нападательным, защитным, территориальным и ритуальным. Лимбическая система отвечает за эмоциональную составляющую жизнедеятельности животного, а также за отдельные виды памяти. Наконец, неокортекс свойствен только высшим животным и отвечает за такие функции, как восприятие и самовосприятие, обучение и память, принятие решений и абстрактное мышление. Предполагается, что именно в неокортексе располагается сознание человека.




Соответствие моделей мозга Аристотеля и Поля МакЛина

Другими словами, только лишь по внешним наблюдениям и при помощи спекулятивных размышлений Аристотелю удалось предвосхитить одну из наиболее интересных моделей мозга, принятых уже в наше время. Впрочем, это действительно можно сделать, наблюдая и сопоставляя поведение различных живых организмов.

Из-за того что у древних мыслителей всё-таки не было методов изучения человеческого мозга во всех его аспектах и функциональности, поискам природы разума посвящалось незначи-

тельное количество усилий. А в темные времена, наступившие в Европе, этот вопрос был совсем отставлен. Потихоньку изучение человеческого интеллекта и его сущности начало выходить на повестку дня только во времена эпохи Возрождения. Но и тогда, несмотря на становление философии гуманизма, философам и мыслителям было сложно подступиться к этой теме.

Прорывом стала книга Рене Декарта «Рассуждения о методе, чтобы хорошо направлять свой разум и отыскивать истину в науках», которую он опубликовал в 1637 году. Несмотря на то что в то ужасное время гонения на свободных разумом учёных было делом обыденным, Декарт отважился на публикацию этой поистине прорывной работы (хотя если говорить откровенно, на современном языке эту работу лучше всего назвать «подрывной»). Она стала знаковой и, по мнению многих историков науки и философии, ознаменовала собой переход к современному научному познанию и к философии Нового времени. Фактически книга «Рассуждения о методе...» стала базисом всей современной эпистемологии.

Декарт провозгласил механистический подход к пониманию природы и объективной реальности, т. е. рассмотрение объектов в окружающем мире в качестве неких механизмов, подчиняющихся физическим законам. Это позволило отвергнуть так называемый витализм, когда движущей силой живой материи объявлялась некоторая сверхъестественная сущность, не подчиняющаяся законам физики, химии, биологии. Собственно с этого момента учёные смогли переключить перспективу и начать рассматривать разум и интеллект человека в той же парадигме.

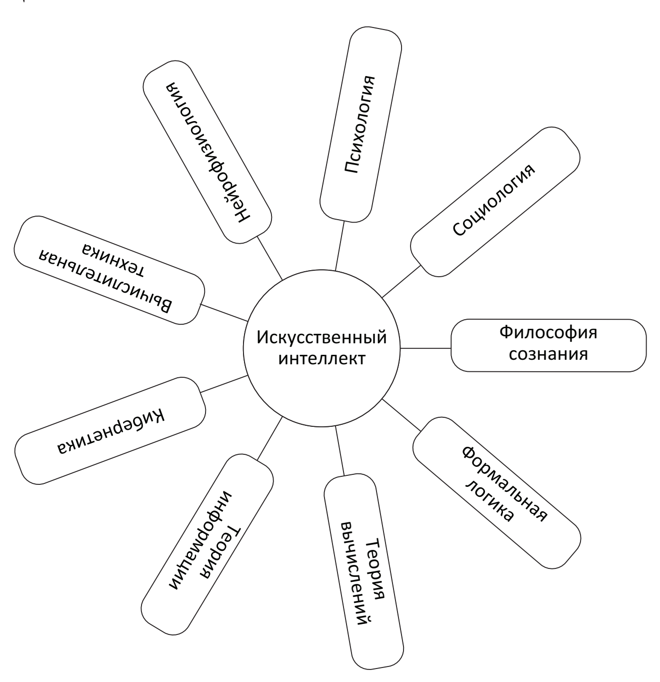
Однако дальше дело не пошло, поскольку, кроме научного метода, у учёных до второй половины XIX века просто не было достаточной инструментальной базы как с точки зрения математических формализмов для исследования феноменологии мышления, так и с точки зрения физических и химических инструментов для исследования материи мозга.

«Взрывное» развитие математики и вслед за ней прикладных наук начиная со второй половины XIX века дало стимул к развитию и различным направлениям в области изучения мышления и интеллекта. Но здесь уже сложно говорить о какой-то последовательной работе над этой темой, так как сам по себе вопрос изучения человеческого разума является междисциплинарным и про историю развития наук об искусственном интеллекте можно говорить исключительно в ключе постепенного сбора информации

и разработки методов в совершенно различных областях знания, которые затем объединялись и в конечном итоге привели к выделению искусственного интеллекта в отдельное научное направление. Другими словами, искусственный интеллект, как дерево, базируется на мощной корневой системе, в которую отдельными корешками входят различные науки от философии, чистой математики и теории вычислений до нейрофизиологии и психологии.

Так что можно выделить следующие магистральные направления научной мысли, которые питают искусственный интеллект (не включая отдельные более узкие специализации):

- философия сознания;
- формальная логика;
- теория вычислений;
- теория информации;
- кибернетика;
- вычислительная техника;
- нейрофизиология;
- психология;
- социология.



Итак, исследования в рамках философии сознания начал Рене Декарт, после которого к этой теме приобщались многочисленные философы, среди которых стоит отметить Джона Локка и Дэвида Юма. Много усилий к изучению природы сознания прилагали философы немецкой школы, в частности Артур Шопенгауэр и Георг Гегель. К XX веку философия сознания выделилась в отдельное направление в рамках *аналитической философии*, которое является чуть ли не единственной прикладной философской дисциплиной, имеющей большое значение для многих направлений научной мысли, в том числе и для искусственного интеллекта. Философия сознания пытается ответить на вопросы о природе сознания и его соотношения с объективной реальностью, что связано с моралью, свободой воли и этическими вопросами, которые, в свою очередь, сразу же возникают при более глубоком изучении проблем искусственного интеллекта и развития его взаимоотношений с интеллектом естественным.

Как уже было упомянуто, основы формальной логики заложил ещё в античные времена Аристотель, однако потом до начала XX века эта методология использовалась учёными без какого-либо развития. Математики пытались как-то формализовать научный метод, и даже было произведено несколько интересных попыток. Немецкие математики и логики Георг Кантор и Готлоб Фреге фактически стали отцами наивной теории множеств и теории предикатов первого порядка соответственно. Эти теории позволяли формализовать очень многое, однако страдали от важного недостатка – противоречивости. И только в 1910–1913 гг. английские математики и философы Бертран Рассел и Альфред Уайтхед опубликовали трёхтомную работу «Принципы математики», где они ввели теорию типов как инструмент для более точной формализации основ математики, в которой невозможно было сформулировать парадокс Рассела о «множестве всех множеств». Именно после этой книги развитие математики пошло семимильными шагами, в результате чего сделали свои открытия Курт Гёдель, АлонзоЧёрч, Алан Тьюринг и многие другие замечательные учёные. Так что в историческом ряду развития формальной логики стоят такие личности, как Аристотель, Г. Кантор, Г. Фреге, Б. Рассел, А. Уайтхед и К. Гёдель.

После того как в 1931 г. австрийский логик Курт Гёдель опубликовал свои работы, в которых было приведено доказательство его знаменитых теорем о неполноте, начались исследования в этом

направлении. Многие из них были чисто философскими, однако две примечательные работы легли в основу всей современной вычислительной техники. Первая – лямбда-исчисление Алонзо-Чёрча, описанное в его теперь уже знаменитой статье 1936 г., в которой он показал существование неразрешимых задач. Параллельно ему Алан Тьюринг переформулировал теорему Гёделя и, пытаясь решить «проблему разрешения» Давида Гильберта, разработал формализм в виде гипотетического устройства, которое впоследствии стало носить название «машины Тьюринга».

Обобщение достижений А. Чёрча и А. Тьюринга привело к формулированию тезиса Чёрча–Тьюринга, который, являясь эвристическим утверждением, гласит, что для любой алгоритмически вычислимой функции существует вычисляющая её значения машина Тьюринга. Этот тезис постулирует эквивалентность между интуитивным понятием алгоритмической вычислимости и строго формализованными понятиями *частично рекурсивной функции* (по Чёрчу), или *функции, вычислимой на машине Тьюринга* (по Тьюрингу). Тезис невозможно строго доказать или опровергнуть ввиду того, что интуитивное понятие алгоритмической вычислимости строго не определено. Однако этот тезис в совокупности с теорией вычислений сегодня лежит в основе алгоритмического решения задач и, как следствие, имеет непосредственное применение в рамках искусственного интеллекта.

Вместе с тем в 1948 году американский математик Клод Шеннон публикует статью «Математическая теория связи», которая сегодня считается вехой в рождении теории информации. Несмотря на то что до К. Шеннона в области проблем передачи информации работали такие пионеры, как Гарри Найквист и Ральф Хартли, именно Клоду Шеннону удалось математически точно сформулировать основные положения новой науки, определить её базис и доказать основную теорему, позже названную его именем. Эта теорема определяет предел максимального сжатия данных и числовое значение информационной энтропии. В связи с дальнейшим развитием сетей передачи данных теория информации и все её приложения стали развиваться семимильными шагами, что привело к появлению многочисленных способов помехоустойчивого кодирования информации для каналов с шумом. Всё это имеет самое непосредственное значение для развития интеллектуальных систем, поскольку вопросы передачи информации в них стоят на одном из первых мест.

Конечно, одной из центральных наук, стоящих в основе разработки технологий искусственного интеллекта, является кибернетика как базовая методология исследования сложных систем, взаимодействующих друг с другом и со средой. Кибернетика сама по себе является междисциплинарной областью исследования, базирующейся на многих отраслях науки, в том числе и уже перечисленных здесь ранее. Однако именно разработанный в её рамках научный аппарат в полной мере позволяет целенаправленно заниматься поиском и проектированием сложных адаптивных и самообучающихся систем, к которым, вне всяких сомнений, относятся системы искусственного интеллекта. Кибернетику как науку разрабатывали такие знаменитые учёные, как Уильям Росс Эшби, Карл Людвиг фон Берталанфи, Джон фон Нейман, Стаффорд Бир, а также многочисленная когорта русских учёных, среди которых обязательно надо упомянуть Ивана Алексеевича Вышеградского (основоположника теории автоматического регулирования), Алексея Андреевича Ляпунова, Виктора Михайловича Глушкова и популяризатора науки об искусственном интеллекте Льва Тимофеевича Кузина.

Все теоретические изыскания в области теории вычислений, теории информации, кибернетики и других наук в конечном итоге приводят к развитию вычислительной техники как прикладной дисциплины, рассматривающей вопросы создания и программирования универсальных вычислительных машин. Попытки создать механическую машину для вычислений предпринимались со времён, наверное, Блёза Паскаля, и первым в этом преуспел наш соотечественник Семён Николаевич Корсаков, который в 1830-х годах создавал первые «интеллектуальные машины» на перфокартах. Хотя в те же самые годы английский математик Чарльз Бэббидж разрабатывал проект универсальной цифровой вычислительной машины, до реализации дело у него не дошло. Несмотря на всё это, основы современной вычислительной техники были заложены Джоном фон Нейманом, который разработал принципы построения архитектуры универсальных вычислительных машин. Впрочем, первый компьютер в современном понимании сделал немецкий инженер Конрад Цузе, он же разработал и первый язык программирования высокого уровня, однако из-за военно-политических особенностей мира в те времена работы Цузе оставались малоизвестными. После Второй мировой войны работы над созданием универсальных компьютеров ве-

лись во всё ускоряющемся ритме, который выдерживается до сих пор (так называемый «закон Мура»).

Однако по мере продвижения в деле создания всё более мощных вычислительных систем становилось ясно, что разработать интеллект *in silico* так просто не получится, и исследователей в этом отношении ждало такое же разочарование, как создателей различных механических кукол (автоматонов), которым казалось, что ещё чуть-чуть – и их куклы обретут разум. Как не получилось с механикой, так же не получилось и с электроникой. Это способствовало вовлечению в исследования нейрофизиологов и других специалистов в части анатомии, физиологии и других аспектов функционирования нервной системы и других регуляторных систем организмов человека и животных.

И вот в 1943 г. американские нейрофизиолог Уоррен Мак-Каллок и математик УолтерПиттс публикуют статью, которая открыла миру новую вычислительную модель, основанную на понятии искусственного нейрона. Да, эта модель была довольно упрощённой и не принимала во внимание большое количество свойств органических нейронов, однако она позволяла производить вычисления. Эта статья фактически открыла широчайшее направление исследований, которое сегодня превалирует в области искусственного интеллекта – искусственные нейронные сети. Вслед за У. МакКалоком и У. Питтсом следует отметить таких учёных, как канадский физиолог Дональд Хебб, который описал принципы обучения искусственного нейрона (он предложил первый работающий алгоритм обучения искусственных нейронных сетей), а также американский нейрофизиолог Фрэнк Розенблатт, который разработал на искусственных нейронах устройство, моделирующее процесс восприятия, – перцептрон.

Но всё, как обычно, оказалось не таким простым, как казалось. Несмотря на то что исследователям удалось смоделировать один нейрон и составить из таких моделей нейронную сеть, сознания в ней так и не зародилось. С одной стороны, это было связано с тем, что на тех вычислительных мощностях, которые были доступны учёным в середине XX века, можно было смоделировать нейронную сеть, состоящую из пары сотен нейронов и нескольких слоёв. Такой объём совсем не соответствует десяткам миллиардов нейронов в головном мозге человека с сотнями тысяч связей для каждого нейрона. С другой стороны, становилось понятно, что «карта не является местностью», так что ждать самоза-

рождения сознания в нейронной сети, даже если она будет очень сложной, слишком странно. Поэтому исследователи обратились к такой науке, как психология.

Одним из первых учёных, кто обратил внимание на этот аспект искусственного интеллекта и его отношения к человеческому разуму, был русский учёный-медик и кибернетик Николай Михайлович Амосов. В ряду его обширной библиографии есть такие знаковые работы, как «Искусственный разум», «Автоматы и разумное поведение» и «Алгоритмы разума». Несмотря на глубокую степень проработки ряда важнейших вопросов, в этих работах всё так же остаётся нераскрытым вопрос о природе сознания. И получается довольно парадоксальная ситуация. Есть практически полное понимание того, как работает нейрон на уровне клетки и субклеточных структур вплоть до биохимических реакций и метаболических путей преобразования веществ, что фактически и эмулирует вычислительные процессы в рамках одной клетки. Также есть понимание, но уже не такое целостное, того, как работают нейронные сети. А ещё есть понимание психологии человека, его поведения и высших когнитивных и интеллектуальных функций. Это понимание ещё менее глубокое, но есть множество операционных гипотез, которые позволяют описывать, объяснять и предсказывать. Но при этом нет никакого понимания того, что находится посередине. Каким образом биохимические реакции нейрона и электрохимические процессы в нейронных сетях приводят к возникновению сознания, интеллекта и разума? Ответа на этот вопрос до сих пор нет. И в итоге получается, что психология – это попытки «дизассемблировать» высшую психологическую деятельность центральной нервной системы человека, но они отвязаны от базовых химических и физических процессов.

Наконец, учёные задумались и о таком важном аспекте интеллектуальной деятельности, как «коллективный интеллект». Само человеческое общество в целом часто показывает более высокий уровень интеллектуальной деятельности, чем каждый его отдельный представитель. Ведь многие сложные научные концепции и технические объекты могут быть изобретены и разработаны исключительно в рамках объединения усилий разноплановых специалистов. Однако не только человеческое общество показывает подобный паттерн поведения. Он виден и в жизни, например, общественных насекомых, когда каждая отдельная

особь, вообще не имеющая и тени разума, делает вклад в поведение своего «суперорганизма», который кажется вполне интеллектуальным. Муравьи и пчёлы – это вершина эволюции насекомых на Земле. Эти наблюдения натолкнули исследователей в области искусственного интеллекта на идеи о «роевой модели интеллекта» в рамках так называемых многоагентных систем. Это – одно из самых быстро развивающихся сегодня направлений науки и техники. Здесь сложно назвать какую-либо ключевую личность, поскольку тема возникла совсем недавно. Но эта тема очень горячая и ещё долго будет оставаться на острие научного поиска.

Итак, мы рассмотрели большую часть научных направлений, которые лежат в основе искусственного интеллекта. Это, можно сказать, его базис и научные и технологические предтечи, предпосылки. Когда же родилось само направление исследований, которое назвали «искусственный интеллект»? Для ответа на этот вопрос необходимо обратиться к работам философов и практиков искусственного интеллекта Марвина Мински и Джона Маккарти. Второй так вообще является автором самого термина «искусственный интеллект» (а кроме того, он разработал язык программирования LISPи является одним из основоположников функционального программирования). Эти учёные основали в 1959 г. Лабораторию информатики и искусственного интеллекта в рамках Массачусетского технологического института, и это была первая научная лаборатория, которая занималась данной проблемой.

Именно Джон Маккарти сформулировал основополагающие принципы искусственного интеллекта, определив то, что потом было названо «чистым подходом», или «нисходящим искусственным интеллектом», и выразилось в гипотезе Ньюэлла–Саймона о том, что осмысленные действия можно выполнять только при наличии в некоторой системе механизма символьных вычислений, а сами такие символьные вычисления являются необходимым условием наличия в этой системе интеллекта. Другими словами, подход Джона Маккарти выражался в том, что системы искусственного интеллекта должны имитировать высокоуровневые психологические процессы разумного существа,

такие как логическое мышление, логический вывод, речь, творчество и т. д.

С другой стороны, его друг и коллега Марвин Мински сформулировал совершенно противоположное определение искусственного интеллекта, которое получило наименование «грязного подхода», или «восходящего искусственного интеллекта». В основе этой парадигмы лежит попытка моделирования естественных процессов, происходящих в самой природе человека. В первую очередь это, конечно же, моделирование нейросетевых процессов в разных аспектах. Наиболее широко проявившейся технологией в рамках грязного подхода стали искусственные нейронные сети, которые моделируют разные процессы человеческого разума на логическом уровне. Можно было бы попробовать смоделировать биохимический уровень, однако для этого не хватает вычислительных мощностей даже сегодня, не говоря уже про те давние времена. Другой известной технологией являются различные генетические и эволюционные методы решения задач. Но в целом этот подход не является «искусственным интеллектом» в том понимании Джона Маккарти, как он определил данный термин.

При этом необходимо понимать, что первоначально предназначением той междисциплинарной области исследований, которая получила название «искусственный интеллект», было моделирование когнитивных функций человека для их исследования на модели, чтобы понять природу интеллекта, разума и сознания человека. Другими словами, искусственный интеллект первоначально рассматривался как довольно фундаментальная область исследований, и только через какое-то время появилась задача по практическому применению наработок, которая нашла своё отражение в создании большого количества искусственных систем, решающих задачи, традиционно относившиеся к прерогативе человека.

Фактически две парадигмы, описанные выше, лежат в основе всякого подхода к разработке искусственного интеллекта. На сегодняшний день таких подходов выделяют семь:

1) интуитивный;
2) логический;
3) символьный;
4) структурный;
5) эволюционный;

6) квазибиологический;
7) агентный.

Надо отметить, что агентный подход к разработке искусственного интеллекта чаще всего рассматривается в рамках так называемой гибридной парадигмы, которая представляет собой смесь нисходящей и восходящей парадигм, берет из них лучшее и старается нивелировать отрицательные стороны. Гибридная парадигма и агентный подход будут рассмотрены в самом конце этой главы.

Интересно то, что в рамках искусственного интеллекта, который, как уже было отмечено, является междисциплинарным научным направлением исследований, имеется ряд задач, которые решаются методами всех или некоторых перечисленных подходов. В частности, к таким задачам традиционно относят:

- поиск информации;
- обработка естественного языка;
- представление знаний;
- машинное обучение;
- распознавание образов;
- интеллектуальный анализ данных, или «дата-майнинг»;
- обработка НЕ-факторов знания;
- принятие решений;
- робототехника;
- роевой интеллект.

Нельзя сказать, что это полный и консистентный список задач, к тому же некоторые из представленных задач в какой-то части пересекаются. Тем не менее это хороший список, являющийся анкером, от которого можно отталкиваться при изучении подходов и методов искусственного интеллекта. Другие исследователи могут предлагать иной список и классификацию задач, но далее в этой книге будет рассматриваться решение именно этих задач разными методами искусственного интеллекта, составляющими те или иные подходы в рамках одной из трёх парадигм.

Что интересно, если попытаться расположить в матрице подходы и парадигмы в строках, а решаемые задачи в столбцах, то получится своеобразная «периодическая система технологий искусственного интеллекта», в ячейках которой будут перечислены различные методы конкретного подхода для решения конкретной задачи. Вот так может выглядеть такая матрица.

| Парадигма | Подход | Поиск | Обработка ЕЯ | Представление знаний | Машинное обучение | Распознавание образов | Дата-майнинг | НЕ-факторы | Принятие решений | Робототехника | Роевой интеллект |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Нисходящая парадигма | Интуитивный подход | | Тест Тьюринга **Tt** | | | | | | | Расширенный **Xt** Тест Тьюринга | |
| Нисходящая парадигма | Логический подход | Поиск **Is** информации | Модель **Mm** Маркова | Продукционная **Pr** модель | Дедуктивное **De** обучение | Сопоставление **Pm** с образцом | Регрессионный **Rg** анализ | Теория **Ds** Демпстера - Шефера | Универсальный **Gs** решатель задач | **Rb** На основе правил | Автоматы **Au** |
| Нисходящая парадигма | Символьный подход | Поиск в **Ss** пространстве состояний | Формальные **Fg** грамматики | Семантические **Sn** сети | Базы знаний **Kb** | Семантическая **Sf** свёртка | Деревья **Dt** решений | Нечёткая **Fl** логика | Экспертные **Es** системы | Фреймы **Fr** | Языки **Li** взаимодействия |
| Гибридная парадигма | Агентный подход | Случайное **Rw** блуждание | | | Обучение с **Rl** подкреплением | | | | Кибернетическая машина **Cm** | Роботы **Ro** | Рациональные **Ra** агенты |
| Восходящая парадигма | Структурный подход | | Стат. методы **Sm** обработки ЕЯ | | Искусственные **An*** нейронные сети | | Стат. методы **Sd** дата-майнинга | | | | |
| Восходящая парадигма | Эволюционный подход | Генетические **Ga** алгоритмы | | | Нейро- **Ne** эволюция | | Эволюционное **Ep** программирование | | | | Искусственная **Al** жизнь |
| Восходящая парадигма | Квазибиологический подход | ДНК- **Dc** компьютер | | | | | | | | Био- **Be** электроника | Наноботы **Nb** |

| * | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| * | Искусственные нейронные сети | Перцептроны **Pc** | Сеть **Wn** Ворда | Машина **Bm** Больцмана | Глубокая **Db** сеть доверия | Карта **Km** Кохонена | Авто- **Ae** кодировщики | Нейронная **Nt** машина Тьюринга | Свёрточные **Fn** нейросети | Генеративно- **Gn** состязательные сети | Глубинное **Dl** обучение |

Периодическая система технологий искусственного интеллекта

Далее в этой книге во второй главе описано большинство представленных методов решения типовых задач искусственного интеллекта. Ну а пока рассмотрим каждый из перечисленных подходов к построению искусственных интеллектуальных систем подробнее.

**Интуитивный подход** к искусственному интеллекту был предложен Аланом Тьюрингом в своей ставшей уже знаменитой статье «Вычислительные машины и разум», опубликованной в 1950 г. В ней он предложил процедуру, которая, по его мнению, должна определить на интуитивном уровне, обладает некоторая система интеллектом или нет. Впоследствии эта процедура получила наименование «тест Тьюринга», подвергалась многочисленной критике, была расширена для систем с богатым набором сенсоров и, в общем, дожила до наших дней в качестве вполне себе операционной процедуры. Тем не менее сам подход не является конструктивным, поскольку Тьюринг нигде не говорил о том, *как* построить интеллектуальную систему.

Честно говоря, сам Тьюринг изложил свой тест немного путано (да ещё и в нескольких вариантах), так что сегодня исследователи ломают копья на тему того, каким образом интеллектуальная система должна проходить данный тест. Но большинство исследователей сходится в одном: прохождение системой теста Тьюринга – необходимый фактор для того, чтобы признать за системой наличие интеллекта, но отнюдь не достаточный. Впрочем, и с этим утверждением многие будут спорить, так как сама природа человеческого интеллекта до сих пор не ясна, потому и бессмысленна разработка какого-либо теста для проверки на наличие интеллекта, «похожего на человеческий». Тем более что основная критика интуитивного подхода заключается в том, что моделирование естественного интеллекта не является единственной возможностью создания интеллектуальной системы. Искусственный интеллект может быть построен на совершенно иных принципах, как это ранее бывало со многими иными изобретениями (колесо для передвижения, реактивная тяга для полёта по воздуху, радиоволны для систем связи и т. д.).

Тем не менее, несмотря на то что тест Тьюринга и интуитивный подход в целом не могут считаться серьёзными инструментами в вопросе разработки искусственного интеллекта, сам по себе тест Тьюринга позволяет определить тот минимальный набор технологий и решаемых задач, которые должна обеспечи-

вать система, чтобы считаться интеллектуальной. Кратко перечислим их.

1. *Обработка естественного языка*: интеллектуальная система должна уметь общаться с человеком на естественном языке, воспринимая все его неоднозначности, неопределённости и умолчания.
2. *Представление знаний*: в рамках искусственного интеллекта должны быть представлены как общие, так и специальные знания, при этом система должна постоянно обучаться и пополнять свою базу знаний, в том числе и в процессе диалога с человеком.
3. *Логический вывод*: используя знания и получаемые на вход запросы от человека, интеллектуальная система должна осуществлять правдоподобный логический вывод, который позволяет сформировать ответ на том же естественном языке.
4. *Машинное обучение*: система искусственного интеллекта должна быть адаптивной и приспосабливаться к меняющейся ситуации в общении, используя имеющиеся у неё знания в качестве шаблонов и применяя их к схожим ситуациям и, само собой разумеется, актуализируя по результатам свои знания об окружающей среде.
5. *Дополнительные сенсоры и исполнительные устройства* используются в так называемом «полном тесте Тьюринга», в котором система искусственного интеллекта должна действовать в естественной среде обитания человека, воспринимая её при помощи таких же датчиков, какие есть у человека (видеокамеры, аудиосенсоры, газоанализаторы и др.), и воздействуя на среду при помощи разного рода манипуляторов.

Перечисленное уже наводит на размышления о том, что тест Тьюринга направлен на выявление того, что искусственный интеллект должен успешно «мимикрировать» под человека, хотя для наличия интеллектуальных способностей это совершенно не требуется. Ни одна из перечисленных технологий сама по себе не необходима для того, чтобы считать искусственный объект интеллектуальным. Но в целом наличие этих пяти пунктов с прохождением полного теста Тьюринга позволяет говорить о том, что система *может* иметь интеллект, похожий на человеческий.

**Логический подход** основывается на формальной логике. Ведь ещё древнегреческий философ Аристотель сделал успешную по-

пытку формального описания законов человеческого мышления. Сложно сказать, насколько эти законы универсальны, ведь кроме человеческого мышления у нас нет никаких иных примеров, но общее осмысление данного вопроса подсказывает, что формальная логика относится к чистому математическому знанию, т. е. находится в мире чистых идей, а потому *может быть* общезначимой. Другими словами, мышление, основанное на формальной логике Аристотеля, может быть универсальным, а потому реализация логических правил в искусственной системе может сделать её интеллектуальной. Однако тут имеется та же самая ловушка, как и в случае интуитивного подхода к искусственному интеллекту. Логическое мышление является необходимым условием, но никак не достаточным. Система, обладающая разумом, будет показывать поведение, подчиняющееся законам формальной логики. Но если какая-либо система действует по этим законам, это совсем не значит, что она интеллектуальна.

В рамках математики разработано большое количество формализмов, описывающих логику. Двоичная логика Аристотеля является базовой, над которой надстроены такие варианты, как многозначная логика Лукасевича, нечёткая логика Заде, бесконечнозначная логика антиномий, интуиционистская логика и некоторые другие. Каждый новый формализм был разработан для того, чтобы учесть какие-либо нюансы человеческого мышления и способа принятия решений человеком. Ведь в процессе исследований в направлении логического подхода становилось понятным, что двоичная логика, хотя и является универсальным инструментом размышления и вывода знаний, плохо справляется с такими простейшими аспектами человеческого поведения, как принятие решений в условиях неопределённости, неполноты знания, неточности измерений и т. д.

Однако логический подход сталкивается с серьёзными трудностями, когда возникает необходимость описания неформальных знаний, которые плохо формализуются. Более того, из-за неполноты нашего понимания природы человеческого мышления логический подход всё так же страдает от невозможности полностью описать процесс мышления и принятия решений. Такие феномены, как озарение, интуитивный поиск решения или эмоциональные влияния на принятие решений, не могут быть описаны в рамках логического подхода, хотя они, вне всяких сомнений, являются одним из компонентов человеческого разума.

Тем не менее логический подход составляет основу упомянутой ранее нисходящей парадигмы искусственного интеллекта.

**Символьный подход** ставит во главу искусственного интеллекта способность человека манипулировать символами (в общем понимании этого термина) при осуществлении своей интеллектуальной деятельности. Тут нужно вспомнить термин «вторая сигнальная система», предложенный русским физиологом Иваном Петровичем Павловым для описания абстрактной системы обозначений, которые используются при мышлении человеком в отрыве от непосредственных ощущений, получаемых всеми сенсорными подсистемами нервной системы. Именно наличие второй сигнальный системы, по мнению некоторых исследователей, отличает разумное существо от животного.

Для упрощения рассмотрим кибернетическую цепочку возникновения реакции человека на стимул из внешней среды на примере зрения. Фотоны различной длины волны попадают на рецепторы сетчатки глаза, где запускают каскад биохимических реакций, результат которых заключается в генерации нервного импульса, идущего по глазному нерву в нервные центры в головном мозге человека. Далее этот импульс диспетчеризуется в ядрах зрительного перекрёста и таламусе, после чего попадает в различные зоны коры головного мозга – первичную зрительную кору, вторичную зрительную кору и т. д., пока след от этого импульса не будет обработан в высших слоях неокортекса, где, как предполагается, и осуществляется манипуляция символами. Именно здесь конкретные каскады биохимических реакций в клетках сетчатки и всех промежуточных нейронах каким-то образом преобразуются в *абстрактные символы*, имеющие огромное количество ассоциативных связей. И когда человек видит, скажем, кошку, то в его памяти возникают различные образы, именно на символьном уровне связанные с понятием «*Feliscatus*» («кот домашний», биологическое наименование вида). Человек осознаёт именно эти образы и ассоциативные связи, но никак не осознаёт лежащих в их основе электрофизических и биохимических процессов. Так что разум человека, его интеллект оперирует именно символом «кот домашний», пробегая по отдельным ассоциативным связям, выбор которых зависит от текущего контекста.

Обработка символов построена на выполнении правил различного вида. В дальнейшем мы рассмотрим несколько формализмов, которые конкретизируют и реализуют символьный подход

к искусственному интеллекту. Однако все они основаны именно на манипулировании символами как синтаксическими конструкциями. Особенно это касается математических формул, поскольку наибольших успехов символьный подход достиг именно в формальной математике, в таких областях, как автоматическое доказательство теорем, символьная математика, автоматические вывод и рассуждения и т. д. Однако и тут возникает такая же возможность для критики, как и в случае интуитивного подхода, – где возникает сознание, в какой момент интеллектуальная система начинает осознавать себя. И как бы хорошо она ни манипулировала символами, ответить на этот вопрос на сегодняшний день возможным не представляется. Мы ещё вернёмся к этому аспекту в главе про философию сознания и искусственного интеллекта, а пока перейдём к следующему подходу.

**Структурный подход**, или *коннекционизм*, исходит из понимания того, что интеллект, разум и сознание являются функцией сложности сети переплетённых и взаимодействующих базовых элементов. Поскольку нервная система человека состоит из нейронов и ряда других специальных клеток, то основой структурного подхода в искусственном интеллекте является использование искусственных нейронных сетей.

Здесь уже упоминалось, что первый искусственный нейрон был разработан Уорреном Мак-Каллоком и Уолтером Питтсом в первой половине XX века. Также в то время Фрэнком Розенблаттом была разработана первая архитектура искусственных нейронных сетей – перцептрон. Фактически это самые первые примеры математических моделей и программных систем, разработанных в духе коннективизма. И здесь необходимо отметить, что в процессе развития перцептронов появились некоторые расширения первоначальной модели, предложенной Ф. Розенблаттом. Самый простой классификатор основан на количестве слоёв в перцептроне: однослойный, с одним скрытым слоем (классический) и многослойный. Все эти типы были в своё время описаны самим Ф. Розенблаттом.

Другая классификация включает: элементарный перцептрон, простой перцептрон, перцептрон с последовательными связями, перцептрон с перекрёстными связями, перцептрон с обратными связями, перцептрон с переменными связями. Первые три класса были описаны Ф. Розенблаттом, а следующие три развиты в дальнейшем при детальной проработке модели искусственных нейронных сетей.

Итак, структурный подход основан на идее о том, что наиболее важной для эмерджентного проявления разума и других подобных функций вплоть до сознания является сетевая структура. Это значит, что моделирование должно затрагивать не только базовые элементы типа нейронов, но и их взаимосвязанные сети. При этом полагается, что сеть элементов как сложная система обладает нелинейными свойствами относительно базовых элементов, а потому при увеличении количества взаимосвязанных элементов сложность модели изменяется нелинейно и зачастую даже непредсказуемым образом, так что при переходе некоторого порога сложности как раз и начинают проявляться те самые эмерджентные эффекты, которых мы так ждём. До сегодняшнего дня главным и фактически единственным представителем структурного подхода являются искусственные нейронные сети.

**Эволюционный подход** реализует так называемую «искусственную эволюцию» и фактически решает оптимизационную задачу поиска значения целевой функции в заданном пространстве решений с установленными ограничениями. Этот подход содержит большое количество эвристических методов, некоторые из которых используются даже для гарантированного нахождения оптимума для мультимодальных недифференцируемых функций в пространствах высоких размерностей. Важным методом эволюционного подхода являются генетические алгоритмы, которые мы рассмотрим чуть позже. Но здесь также есть несколько других важных методов, в числе которых находятся эволюционное программирование и даже нейроэволюция, когда искусственному отбору подвергаются нейронные сети.

Итак, в рамках эволюционного подхода обычно выделяют следующие технологии и методы: эволюционное программирование, генетическое программирование, эволюционные стратегии, генетические алгоритмы, дифференциальная эволюция и нейроэволюция. Другими словами, а что, если вычислительные процессы могли бы эволюционировать так же, как это делают биологические виды в своей экологической среде? Возможно, получилось бы «выращивать» программы для оптимального решения поставленной задачи? Эволюционное программирование как раз и основано на этой идее.

Если же в качестве объектов отбора выступают сами программы, то получается уже генетическое программирование. Ведь, действительно, программы пишутся на определённом языке про-

граммирования и в конечном итоге представляют собой строки символов. Эти строки можно подвергнуть генетическим преобразованиям и отбору. Эта очень мощная идея получила своё развитие в том, что программы начали писать и оптимизировать другие программы, и уже исследователи, запустившие процесс, не могут разобрать и интерпретировать полученные исходные коды, которые работают правильно и часто очень эффективно.

Генетические алгоритмы – это наиболее яркий представитель эволюционного подхода. Сами по себе они опять являются одним из эвристических методов оптимизации для поиска оптимального решения (или, как минимум, субоптимального). Они работают с данными, которые могут быть представлены в виде «хромосом» – последовательностей генов, т. е. списков каких-либо значений, к которым можно применить генетические операции. Здесь главное – чтобы на генах были определены эти самые генетические операции, которые возвращали бы приемлемый результат, имеющий смысл.

Мы ещё детально ознакомимся с эволюционными алгоритмами и вообще эволюционным подходом во второй главе.

Идём дальше и кратко коснёмся **квазибиологического подхода** и его методов. Фактически это отдельное направление исследований в искусственном интеллекте, так как оно основано не на цифровом моделировании *in silico* разных аспектов интеллекта, а на применении биомолекулярных механизмов для того же самого. Этот подход ещё называется «биокомпьютингом», и это очень перспективное направление. В рамках квазибиологического подхода разработано большое количестве методов – начиная от биомолекулярной электроники, молекулярных вычислений и заканчивая нейрокомпьютингом. В последнем направлении важной вехой является разработка нейроморфных чипов. И иногда мне кажется, что прорыв в области искусственного интеллекта будет лежать на пересечении трёх областей – техники, информатики и химии. Но посмотрим…

В основе этого подхода лежит понимание, что феномены человеческого поведения, наша способность к обучению и адаптации есть следствие именно биологической структуры и особенностей её функционирования. Хотя, скорее всего, это очень слабая гипотеза. Вычисления в рамках квазибиологического подхода организуются при помощи живых тканей, клеток, вирусов и различных биомолекул. Часто используются молекулы дезоксирибонуклеи-

новой кислоты, на основе которой создают ДНК-компьютер. Кроме ДНК, в качестве биопроцессора могут использоваться также белковые молекулы и биологические мембраны.

Обычно для решения определённой задачи создаётся так называемая «индивидуальная машина», которая, в отличие от универсальной машины Тьюринга, направлена на решение конкретной задачи, причём обычно делает это более эффективным способом, поскольку индивидуальная машина специально сконструирована для решения именно этой задачи. Машина Тьюринга, лежащая в основе стандартной вычислительной модели, выполняет свои команды последовательно, а в рамках квазибиологической парадигмы часто рассматривается массовый параллелизм. Ну вот если, к примеру, рассмотреть ДНК-компьютер, то в нём все молекулы ДНК одновременно участвуют во взаимодействиях, параллельно проводя вычисления.

Два самых главных направления в рамках квазибиологического подхода – это молекулярные вычисления и биомолекулярная электроника. Можно ещё упомянуть нейрокомпьютинг и создание нейроморфных чипов, но они чаще всего рассматриваются как часть структурного подхода и искусственных нейронных сетей.

Молекулярные вычисления – это отдельная вычислительная модель, в которой решение задачи осуществляется при помощи проведения сложных биохимических или нанотехнологических реакций. Молекулярные компьютеры – это молекулы, запрограммированные на нужные свойства и поведение, которые, участвуя в химических реакциях, как бы «выращивают» результат. Что интересно, идею биокомпьютинга подсказал выдающийся математик Джон фон Нейман в своей книге «Теория самовоспроизводящихся автоматов», которую, кстати, очень рекомендую для внимательного чтения. В этой книге описан проект клеточных автоматов, которые могут самовоспроизводиться, как живая клетка.

Почти в каждой живой клетке нашего организма есть длинная молекула ДНК, кодирующая генетическую информацию. При помощи различных ферментов цепочки ДНК могут быть разрезаны, склеены, в них могут добавляться буквы генетического кода или удаляться из них. Всё это – базовые операции работы с информацией, которые могут быть использованы для производства вычислений. Более того, цепочки ДНК могут воспроизводиться и клонироваться. Это позволяет запустить массовый параллелизм

поиска решения. В небольшой пробирке после проведения должным образом сконструированной биохимической реакции будет получен результат, который считывается специальной аппаратурой.

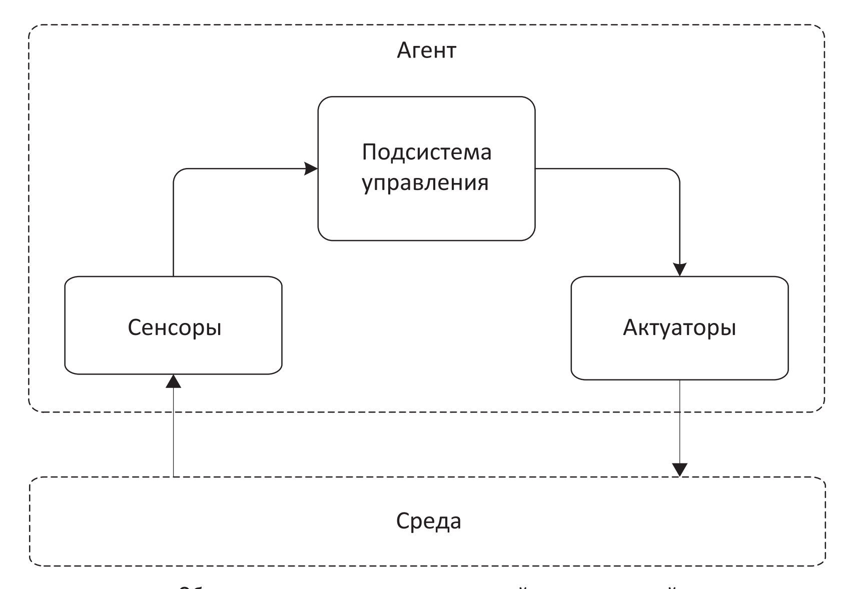
Интерес вызывает то, что для некоторых задач молекулярные компьютеры очень быстро и точно находят приемлемые решения, в то время как традиционные компьютеры затрудняются это сделать. Например, решение задачи коммивояжёра, т. е. поиска кратчайшего пути обхода графа, при помощи реакций с ДНК осуществляется практически мгновенно, в то время как для обычного компьютера требуется огромное количество времени. Правда, тут есть одна тонкость, которая мешает работе обычному компьютеру, – это комбинаторный взрыв. И если в традиционной архитектуре он ведёт к увеличению времени решения, то для ДНК-компьютера требуется подготовка огромного количества вариантов нуклеотидных нитей. Соответственно, объём пробирки растёт так же, как и количество вариантов в комбинаторном взрыве.

В общем, часто биокомпьютинг можно охарактеризовать как новую парадигму вычислений, которая, в отличие от традиционной вычислительной модели, работает быстро, но при решении сложных задач с комбинаторным взрывом растёт не время вычислений, а необходимый для них объём биокомпьютера.

Вместе с тем в последнее время всё активнее разрабатывается **агентный подход** к построению искусственного интеллекта. В рамках этого подхода изменена точка зрения на цель построения интеллектуальной системы и считается, что построить нужно систему не с разумным поведением, а с рациональным. С одной стороны, это серьёзно облегчает задачу, поскольку, в отличие от понятий «разум» или «интеллект», понятия «рациональность» и «рациональное поведение» можно строго формализовать (например, рациональное поведение – это выбор и достижение оптимальной цели с минимизацией затраченных на это ресурсов). С другой стороны, для демонстрации рационального поведения агент должен обладать достаточной «разумностью», чтобы определить цель, составить стратегию её достижения и выполнить её.

Каждый агент – это полноценная кибернетическая машина, которая имеет систему управления, непрерывно получающую информацию с сенсорных систем агента и воздействующую на окружающую среду при помощи исполнительных устройств

(или *актуаторов*). При этом подход не определяет сущность сенсорных систем и актуаторов – их природа может быть произвольной. Поэтому агентный подход одинаково применим как к чисто программным сущностям, работающим в некоторой искусственной среде, так и к программно-аппаратным комплексам, равно как и вообще к биологическим системам.




Общая схема агента и его взаимодействия со средой

Агентный подход интересен тем, что в его рамках можно использовать эволюционные алгоритмы, которые подбирают интеллектуальных агентов, исходя из степени их приспособленности к достижению цели. Во время взаимодействия агентов осуществляется отбор наиболее успешных, которые затем используются для генерации нового поколения агентов, среди которых опять применяются те же самые процедуры оценки и отбора. В итоге наиболее успешное поколение решает задачи и достигает целей наиболее эффективным образом. Это идеальный вариант, который сегодня сложно достижим, но стремиться к нему интересно.

Также агентный подход лежит в основе так называемых многоагентных систем, в рамках которых осуществляется общее целеполагание, после чего каждому индивидуальному агенту даётся свобода действий в определённых рамках, где он имеет возмож-

ности и альтернативы по разработке и реализации различных стратегий достижения своей частной цели. В процессе этого агенты взаимодействуют друг с другом и со средой, обмениваясь информацией и выполняя запросы других агентов. Кроме того, вполне может быть использована идея так называемого «роевого интеллекта», когда каждая отдельная «особь» (то есть агент) интеллектом не обладает, но в целом «рой» (множество агентов, многоагентная система) обладает определёнными интеллектуальными способностями. Надо отметить, что всё перечисленное является одним из наиболее перспективных направлений исследований по искусственному интеллекту.

Итак, философия сознания, начавшаяся развиваться вместе с более «взрослым» пониманием задач и проблем искусственного интеллекта, выделяет два крупных подхода к построению интеллектуальных систем и искусственных разумных существ: чистый (нисходящий) и грязный (восходящий). Чистый подход объединяет такие технологии, как экспертные системы, универсальные машины вывода, семиотические базы знаний. Самым выдающимся и далеко продвинувшимся направлением в рамках этого подхода является направление символьных вычислений, которое основано на логике манипулирования символами. Грязный подход объединяет такие технологии, как искусственные нейронные сети, эволюционные вычисления и биокомпьютинг. Грязный подход моделирует биологические основы разума у человека, в то время как чистый подход имитирует высокоуровневые психические и когнитивные процессы: мышление, рассуждение, речь, эмоции, творчество и т. д.

Однако, как уже было описано выше, и чистый, и грязный подходы имеют как достоинства, так и свои недостатки. В частности, нейронные сети и все смежные технологии обладают очень существенными ограничениями.

1. Наука до сих пор не обладает полным пониманием того, как обучается нейронная сеть. Теоретический математический аппарат, конечно, имеется, однако при переходе в практическое русло вычислительная сложность интерпретации того, как и, главное, почему нейросеть настроила свои весовые коэффициенты так, а не иначе, нелинейно

возрастает до космических величин. И в итоге получается чёрный ящик в качестве модели чёрного ящика. Это совсем не то, что нужно.

2. Часто в нейронную сеть для обучения попадают некорректные данные, которые выглядят как валидные. Например, в медицинской проблемной области нередко случается так, что диагнозы ставятся неверно, лекарственные средства назначаются неадекватно, а потому нет никакой гарантии того, что при обучении нейросети ей не «скармливаются» такие некорректные данные. Принцип GIGO никто не отменял. Кто даст гарантию того, что обученная на некорректных данных нейросеть даст правильное заключение, а инструмента интерпретации того, как она обучилась, нет.
3. В процессе развития человеческих знаний всегда появляется новая информация, которая часто не дополняет, а заменяет старую. Человек-эксперт постоянно актуализирует свои знания, занимаясь самообразованием и повышением квалификации. Как поступать с нейросетью? Ведь невозможно проинтерпретировать то, как она обучилась, а потому нет никакого способа точечно заменить устаревшую информацию. Сеть приходится переобучать с нуля.
4. Более того, «углеродная» нейросеть фундаментально отличается от «кремниевой» – сегодня ещё не разработано окончательной рабочей модели того, как обучается человек. С учётом непрекращающегося нейрогенеза в отдельных областях головного мозга есть понимание, что в нейросетях в голове человека могут появляться не только новые отдельные нейроны, но и даже целые слои. Впрочем, и нейроны, и слои также могут исключаться из процесса вычислений. И если появление и исключение синаптических контактов можно эмулировать весовыми коэффициентами в искусственных нейронных сетях, то про искусственные модели с появлением новых слоёв пока слышно очень редко и тихо.
5. К тому же центральная нервная система человека представляет собой не обычную многослойную нейросеть, а десятки отдельных сетей, связанных друг с другом. Иногда связанных совсем нелинейным образом – с кучей реципрокных и обратных связей, которые могут выходить и входить в промежуточные слои любой сети. Сигнал от сенсора до исполнительного устройства в человеке может пройти через

десяток коммутирующих ядер, запустить параллельные процессы и в итоге так и не дойти до неокортекса, оставшись в подсознательной области.

6. Наконец, что касается неокортекса, где, как предполагается, и скрывается тайна сознания. Обработка информации в нейросети неокортекса может завершиться в любом из шести его слоёв, но только, как считают некоторые исследователи, если сигнал дойдёт до пятого или шестого слоя, он будет доступен для осознания при помощи символьной логики. Вся подсознательная деятельность (можно сказать, что она лежит в основе интуиции) готовит базу для принятия решения на основе механизма манипулирования символами. Именно этот аспект надо рассмотреть более подробно.

Итак, вся накопленная человечеством информация представлена в символьном виде. И именно логическая (синтаксическая) манипуляция символами позволяет делать какие-либо умозаключения. Это совершенно иной подход, фундаментально отличающийся от нейронных сетей, какими бы глубокими они не были. Вполне вероятно, что где-то там в очень глубокой нейросети может зародиться логика манипулирования символами, но пока это неизвестно в силу уже неоднократно упомянутого отсутствия формализма интерпретации работы нейросетей. А с другой стороны, математика предоставляет достаточное число формализмов и инструментов для логической обработки символов.

Символьный подход в искусственном интеллекте сегодня незаслуженно отставлен в сторону. Скорее всего, это произошло из-за раннего разочарования в нём по причине того, что естественный язык как наиболее развитая символьная система, используемая человеком, является очень сложной и плохо формализуемой. Неопределённость высказываний, нечёткость формулировок, разного рода неточность измерений параметров – всё это вносит вклад в повышение сложности принятия решений на основе работы с символьной информацией. И если человек как-то справляется с этим в условиях неопределённости, неполноты и даже противоречивости входной информации, то обучить этому искусственную интеллектуальную систему достаточно сложно. Впрочем, это возможно.

Основная проблема символьного подхода заключается в быстро достигаемом «комбинаторном взрыве» при попытках разобраться в семантике обрабатываемых формул. Большое количество си-

нонимов, нечёткие лингвистические переменные с не до конца определёнными шкалами – всё это отпугивало экспертов и повергало в уныние инженеров. Для того чтобы создать базу знаний для самой простой проблемной области, приходилось либо записывать тысячи правил на все случаи жизни, либо постоянно повышать уровень абстракции представления до того уровня, когда это представление становилось неадекватным решаемым задачам. В качестве примера можно привести базу знаний по выдаче предположений о возможных заболеваниях по общему анализу крови – база содержит более 1000 продукций, но при этом всё ещё не может охватить все возможные случаи, встречаемые в практике.

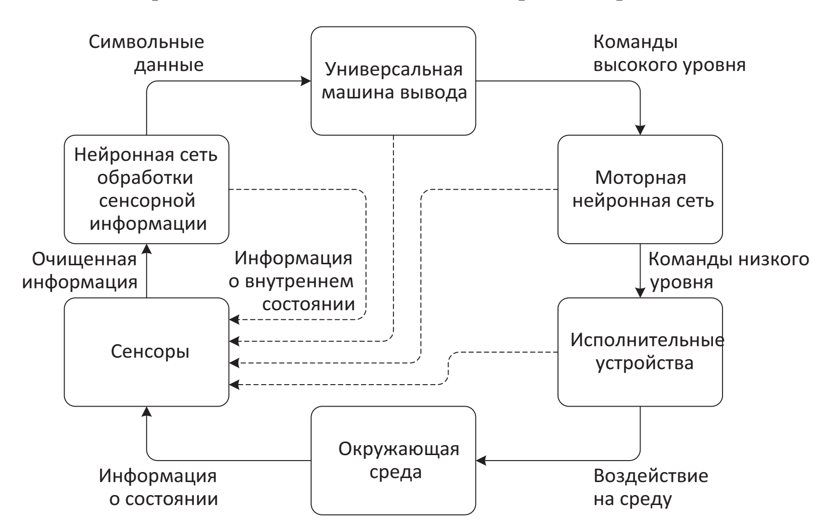
Однако сегодня описанные проблемы вполне поддаются решению. Из-за того же роста вычислительных мощностей, который помог развитию структурного подхода, вполне можно возобновить работы в рамках символьного направления искусственного интеллекта. Во многих проблемных областях знания в подавляющем большинстве случаев либо формализованы, либо хорошо поддаются формализации. Главной задачей становится создание нормальной машины вывода, которая справится с «комбинаторным взрывом».

Уже описанный символьный подход в искусственном интеллекте позволяет решить главную проблему, возникающую при использовании искусственных нейронных сетей и эволюционных алгоритмов, – интерпретацию результатов. Структурированные и формализованные знания и вывод на них позволяют осуществить пошаговое объяснение того, как был получен тот или иной результат. Все остальные особенности структурного подхода вполне доступны и в рамках символьного.

И тут хотелось бы отметить, что рождение искусственного интеллекта, скорее всего, произойдёт где-то на стыке двух подходов, т. е. при использовании **гибридной парадигмы**. Вряд ли в ближайшее время удастся реализовать *in silico* аналог человеческой нервной системы, поскольку её сложность невообразима. Вероятно, проще вырастить биологическую нейросеть такого объёма, чем найти вычислительные мощности для её симуляции. Однако нейросетевой подход даст базовые инструменты для первичной обработки и коммутации информации, поступаемой с разнообразных сенсоров на вход системе управления и принятия решений в составе искусственного интеллекта. А в ней уже будет использоваться символьный подход, реализованный в виде уни-

версальной машины вывода. И такая конвергенция двух технологий в конце концов позволит реализовать давнюю мечту человека – рождение искусственного разума.

Следующая диаграмма графически показывает общую схему взаимодействия компонентов в гибридной интеллектуальной системе, которая основана на описанных ранее принципах.




Как видно из представленной диаграммы, гибридная искусственная интеллектуальная система представляет собой не что иное, как универсальную кибернетическую машину, которая имеет три основных элемента: аффекторы, подсистему управления и эффекторы. При помощи аффекторов кибернетическая машина воспринимает сигналы окружающей среды, которые обрабатываются в подсистеме управления, сигналы из которой далее поступают в эффекторы, воздействующие на окружающую среду. Это общая схема любого автономного агента, поэтому гибридная искусственная интеллектуальная система одновременно является и интеллектуальным агентом, реализуя уже описанный агентный подход.

Однако представленная схема не совсем корректно отражает то, как устроена система принятия решений и реагирования на внешние и внутренние стимулы у человека. Если бы каждый внешний стимул передавался в самую верхнюю систему управления, которая в нашем случае связана с сознательной обработкой информации, то жить и функционировать было бы очень непросто. Даже

предварительная фильтрация незначимых стимулов не спасла бы. Так что систему управления желательно разделить на две подсистемы, и тогда общая схема гибридного интеллектуального агента будет выглядеть так, как показано на следующей диаграмме.




Итак, универсальная машина вывода, или подсистема управления, теперь разбита на две части.

1. *Реактивная подсистема управления* фактически реализует традиционную схему управления, когда сигналы с сенсоров обрабатываются системой управления и по ним осуществляется формирование управленческих воздействий на среду (объект управления) через исполнительные устройства.
2. *Проактивная подсистема управления* добавляет дополнительный промежуточный уровень, который позволяет осуществлять обучение системы, построение прогноза на основе моделирования среды и своего поведения в ней, построение плана действий и сравнение факта с прогнозом и планом для осуществления обучения (адаптации) системы к изменяющимся условиям внешней среды.

Эти подсистемы связаны друг с другом при помощи передачи *фокуса управления*. Когда проактивная подсистема создаёт новый паттерн поведения системы в изменившихся условиях и среда «устоялась», контур управленческого воздействия спускается в реактивную систему, происходит автоматизация реакции, так как фактически для неизменяющихся условий деятельности обучение и проактивное поведение не требуется, а потому реактивная реакция будет работать быстрее. Если же в процессе «рефлекторного» действия реактивной системы обнаруживается изменение в среде или объекте управления, то реактивная система эскалирует фокус внимания на проактивную для обработки изменившихся условий и выработки новых правил и паттернов поведения.

Цикл управления в рамках такой интеллектуальной системы управления теперь заключается в последовательном выполнении следующих шагов.

1. Сбор входной информации со всех сенсоров, которые осуществляют мониторинг различных параметров объекта управления и среды, в которой функционирует система. Каждый тип сенсора в этом случае является отдельной индивидуальной *модальностью восприятия* гибридной искусственной интеллектуальной системы.
2. Сенсоры очищают входную информацию от шумов и осуществляют первый выбор пути дальнейшей обработки. Если входная информация с сенсоров соответствует каким-либо автоматическим *паттернам поведения* системы, то фокус управления передаётся в реактивную подсистему, которая выбирает конкретный паттерн и исполняет его. Однако если в процессе реагировать по реактивному сценарию, система обнаруживает, что что-то пошло не так, осуществляется эскалация на проактивную подсистему управления, как было описано ранее.
3. Если входная информация не имеет автоматической реакции для своей обработки, то осуществляется интеграция всех модальностей восприятия системы в единый блок описания объекта управления и среды. На выходе этого модуля появляется целостная картина восприятия, которая передаётся в проактивную подсистему управления.
4. Проактивная подсистема управления принимает решение на основе имеющихся у неё динамических моделей себя самой, объекта управления и среды. Здесь как раз и использу-

ются методы машинного обучения и нисходящей парадигмы искусственного интеллекта. На выходе проактивной подсистемы управления появляется управленческое воздействие, которое записывается в реактивную подсистему в качестве нового правила, а также направляется на исполнение.

5. Для исполнения управленческое воздействие переводится на язык конкретных исполнительных устройств, которые взаимодействуют с объектом управления и средой. Исполнительные устройства выполняют команду. Цикл работы завершается.

Таким образом видно, что гибридная интеллектуальная система отличается тем, что её аффекторы (сенсоры, датчики) и эффекторы (исполнительные устройства) связаны с подсистемой управления и принятия решений через нейронные сети. Тем самым реализуется «грязный» подход. Аффекторная нейронная сеть принимает очищенные сенсорами сигналы внешней среды и преобразует их в символы, которые подаются на вход универсальной машине вывода. Последняя осуществляет вывод на основе символьных знаний из своей базы знаний и выводит результат, который тоже представляется в виде символов. Тем самым реализуется «чистый» подход. Символьный результат подаётся на вход моторной нейронной сети, которая преобразует высокоуровневые символы в конкретные сигналы управления исполнительными устройствами.

Кроме всего прочего, внутри гибридной интеллектуальной системы должны быть реализованы контрольные связи от всех её элементов к сенсорам. Тем самым реализуются адаптационные механизмы, основанные на гомеостазе внутреннего состояния системы. Сенсоры фиксируют изменение внутреннего состояния каждой подсистемы, их элементов и комплексов, а в случае выхода контролируемых значений за пределы установленных гомеостатических интервалов подсистемой управления принимается решение, целью которого будет возврат изменённых показателей в установочный интервал.

Наконец, связь от проактивной подсистемы управления в саму себя олицетворяет так называемый «внутренний конфликт», когда искусственная интеллектуальная система может рассматривать и моделировать различные варианты развития ситуации. Цикл оценки и выбора приемлемой альтернативы осуществляется до тех пор, пока этот внутренний конфликт не будет улажен.

Интерес вызывает то, что система именно с такой архитектурой при переходе через определённый порог сложности может считаться разумной. Действительно, в соответствии с определением, данным нейрофизиологом Е. Р. Джоном (цитируется по книге Хофштадтера и Деннета «Глаз разума»), сознание – это «процесс, в ходе которого информация о множественных индивидуальных модальностях восприятия и ощущения сводится в единое многоплановое представление о состоянии системы и её окружения и интегрируется с информацией о воспоминаниях и потребностях организма, порождая эмоциональные реакции и программы поведения, способствующие приспособлению организма к его окружению». Представленная на предыдущей диаграмме кибернетическая схема на довольно высоком уровне абстракции описывает, как функционирует нервная система человека и высших животных. Ведь «разумность» в этом смысле определяется как адекватное реагирование не только на стимулы внешней среды, но и на внутренние состояния, что также включает в себя постоянный мониторинг состояния собственной подсистемы управления, что называется «саморефлексией», которая и приводит к осознанию.

Чтобы более точно понять изложенное, имеет смысл рассмотреть несколько примеров. Самым тривиальным примером интеллектуальной системы является человек. На определённом уровне абстракции человек как система выглядит именно так, как представлено на диаграммах выше. Роль сенсоров выполняют не только такие общеизвестные рецепторные системы, как зрение, слух, вкус, обоняние, осязание и равновесие, но и различного рода ноцицепторы, терморецепторы, проприорецепторы и хеморецепторы. Именно последние из перечисленных осуществляют мониторинг гомеостаза внутренних параметров организма, начиная от кислотности крови и заканчивая наличием и концентрацией в физиологических жидкостях организма определённых веществ. Каждый тип рецепторов непрерывно воспринимает сигналы из внешней по отношению к нему среды, при этом сразу же осуществляется первичная фильтрация входных значений по полосе принимаемого диапазона. Например, фоторецепторы на сетчатке глаза воспринимают фотоны только определённых длин волн.

Далее отфильтрованные сигналы «оцифровываются» и передаются в первичные сенсорные нейронные сети, где осуществляется перекоммутация и задействование следующих областей нервной системы. Потом сигнал идёт во всевозможные центры принятия решений, и часто он даже не доходит до некортекса, «поглощаясь» в подсознательных областях (но там также накапливается информация, которая в какой-то момент может стать основой и источником интуитивного инсайта). Если же сигнал доходит до неокортекса, в нём он начинает активно преобразовываться и задействовать множество ассоциативных связей, возбуждая побочные цепи. В итоге до верхних слоёв неокортекса префронтальных долей, где, как предполагается, и находится сознание, доходит символьная информация.

Например, если последовательность множества сигналов, падающих на сетчатку глаза, формирует на ней изображение движущейся рыжей кошки, то на верхних слоях неокортекса активированы те нейроны, которые отвечают за такие символы, как «кошка», «млекопитающее», «рыжее животное» и множество других.

Путём логических умозаключений (на самом деле последовательного возбуждения каскадов нейронов в ассоциативно-теменном и префронтальном участках неокортекса) формируется ответная реакция. Она также кодируется в виде символов. Например, человек, видящий кошку, может выбрать из всего множества ассоциаций одно слово «кошка», сосредоточившись на нём и подав сигнал нижележащим подсистемам активировать поведенческую реакцию типа «воскликнуть наименование объекта». Моторная нейронная сеть воспринимает этот символьный сигнал и начинает распространять его всё ниже и ниже. На самом нижнем уровне символ «кошка» преобразуется в динамическую последовательность сокращений мышц в разных отделах тела, чья совместная деятельность заключается в таком синхронизированном выполнении ряда элементарных сокращений, которые приведут к возникновению исходящего из лёгких потока воздуха с одновременным формированием в голосовых связках, нёбе, языке и зубах быстрого ряда сжатий и растяжений, приводящих к возникновению вибраций определённой амплитуды и частоты. Которые, в свою очередь, воспринимаются окружающими людьми и самим человеком, произнёсшим слово, как слово «кошка». Сам человек воспринимает ушами сказанное самим собой слово, которой через такой же каскад реакций и перекоммутаций

нейронных сигналов доходит до контролирующего элемента неокортекса, что сравнивает заказанное и исполненное слово – они должны совпасть. Эта петля обратной связи контроля поведения широко используется в большинстве поведенческих актов человека. Всё это происходит практически мгновенно и без задействования сознания и внимания. Но похоже, что и сознание основано именно на этом механизме.

Другим примером является интеллектуальная система управления перекрёстком на улично-дорожной сети города. Здесь окружающей средой, в которой действует интеллектуальный агент, является дорожное движение. Сенсорами являются детекторы транспорта и другие периферийные устройства, которые снимают показатели транспортных и пассажиропотоков, значения выделенных погодных показателей и показателей локальной экологии, а также коммуникационное оборудование, принимающее сигналы операторов из центра управления. Кроме того, система постоянно опрашивает все подключённые к ней устройства на предмет их работоспособности, и это функция контроля собственного гомеостаза. Исполнительными устройствами являются светофоры, табло отображения информации и иные средства прямого и косвенного управления дорожным движением. В качестве подсистемы управления выступает контроллер светофорного объекта, выполняющий сценарии управления при возникновении тех или иных шаблонных или особых ситуаций на дороге, характеризующихся выделенными множествами значений входных параметров, включаются те или иные режимы управления дорожным движением.

Сенсорная и моторная нейронные сети здесь редуцированы до декодеров сигналов с оборудования и преобразователей сигналов верхнего уровня из подсистемы управления в замыкания конкретных электрических цепей, зажигающих лампочки или светодиоды в светофорах или на табло. Отличительной особенностью такого интеллектуального агента является то, что он может действовать автономно, а также он может взаимодействовать с другими агентами, входящими в состав интеллектуальной транспортной системы. При этом подобное взаимодействие направлено на кооперативное достижение единой цели оптимизации дорожного движения в целом путём «торга» во имя достижения локальных целей каждого конкретного интеллектуального агента.

ꕤ

Размышления над тем, может ли искусственная система мыслить так, как человек (что бы это ни значило – мы до сих пор не уверены относительно природы сознания), приводят к двум различным пониманиям искусственного интеллекта как практического направления:

- слабый искусственный интеллект;
- сильный искусственный интеллект.

Споры о том, что такое искусственный интеллект, начались практически сразу же после введения в научную практику определения нового понятия. Но до сих пор ответа на вопрос о том, что же такое «искусственный интеллект» и какие критерии можно использовать для того, чтобы определить, является ли искусственная система интеллектуальной и, шире, разумной, нет. С точки зрения философии сознания ответить на этот вопрос не так-то просто. Как уже было рассказано выше, первым разработать хотя бы какие-то операционные критерии попытался Алан Тьюринг в рамках интуитивного подхода. Попытка не очень удачная, тем не менее она вполне применима, хотя в дальнейшем были определены совершенно иные принципы построения интеллектуальных систем. И системы, построенные на этих принципах, вообще не могут быть подвергнуты тесту Тьюринга, хотя по иным критериям должны считаться интеллектуальными.

Масла в огонь подлил один из интереснейших философов современности Джон Сёрль, который предложил разделить понимание слабого и сильного искусственного интеллекта. Под сильным он понимал такой искусственный интеллект, который, если говорить самыми простыми словами, воспринимает самого себя в качестве отдельной личности. Это очень узкое понимание, поскольку привлекает антропоморфное сравнение. Однако искусственное разумное существо не обязательно должно мыслить как человек. Поэтому говорить о том, что оно будет оперировать человеческими категориями, в том числе и при осознании себя, очень опасно с методологической точки зрения.

Но в итоге исследователи и разработчики интеллектуальных систем пришли к соглашению о том, что считать искусственным интеллектом. Под **слабым искусственным интеллектом** начали понимать интеллектуальные системы, которые решают слабо алгоритмизируемые или вовсе неалгоритмизируемые задачи

методами, схожими с теми, которые использует для решения таких задач человек. Другими словами, слабый искусственный интеллект представляет собой реализацию отдельных когнитивных функций человеческого разума для достижения конкретных целей. При этом о наличии у таких систем самосознания, не говоря уже об иных высших функциях психологической деятельности человека, даже не говорится – их нет, и их существование даже не предполагается.

С другой стороны, **сильный искусственный интеллект** подразумевает появление у искусственной интеллектуальной системы самосознания с дальнейшим выходом в автономный режим жизнедеятельности и получением возможности перепрограммировать саму себя с целью усовершенствования. Именно с этим типом связывается точка достижения технологической сингулярности – сильный искусственный интеллект будет экспоненциально развиваться, поглощая всю накопленную человечеством информацию и используя её для собственного роста. Более того, он сможет генерировать новые идеи и концепции, как это делает человек, что в конце концов приведёт к тому, что сильный искусственный интеллект превзойдёт в интеллектуальных способностях всё человечество вместе взятое. Для обозначения таких искусственных разумных существ далее будет использоваться слово «ИскИн».

Наконец, в последнее время возникла новая парадигма в науке о сознании, которая пока ещё остаётся больше фантастической, нежели реальной. Речь идёт о так называемой **гибридизации сознания**, когда человеческий естественный интеллект сливается с искусственным интеллектом и при этом усиливается его методами и технологиями. Фактически цель этого направления исследований состоит в отыскании путей и возможностей для осуществления цифрового бессмертия человеческой личности, т. е. записи её бренного биологического тела на цифровые носители. Это очень занятное направление исследований, которое находится в самом начале своего пути. Но тут есть многое, о чём можно подумать.

※

Необходимо отметить, что в конечном итоге работы над искусственным интеллектом в понимании искусственного разумного

существа могут привести к его появлению. Сильный ИскИн может появиться, так как этому нет фундаментальных преград. Действительно, мы сами, люди, – отчётливый пример того, что разум, интеллект и сознание существуют в этой реальности, а это значит, что все понятия, описывающие конкретные явления, не являются умозрительными. Интеллект так или иначе существует, так что почему же надо отказывать себе в возможности его моделирования или даже создания? Абсолютно не стоит.

Однако размышления наводят на мысли о том, что, скорее всего, ИскИн может появиться в виде гибридной ИИ-системы на стыке, как минимум, трёх дисциплин: информатики, химии с биологией и техники.




Действительно, описанная ранее квазибиологическая парадигма искусственного интеллекта сегодня является одним из мощно развивающихся направлений исследований. В чашках Петри уже научились выращивать искусственные нейронные сети на живых нейронах, которые обучаются управлять простыми механизмами. Бионические протезы и импланты, в том числе нейроимпланты, становятся реальностью на наших глазах. Достижения робототехники поражают сегодня даже видавших виды инженеров и учёных.

Другими словами, информатика и весь комплекс наук в её основе даст понимание основ вычислительных процессов и того, как они моделируют психологические функции. Техника и все

смежные дисциплины дадут возможность создавать физические тела для ИИ-систем. А химия и биология подведут к пониманию того, как в ИИ-системы встроить механизмы метаболизма и эволюции, что в конечном итоге даст ключ к пониманию жизни как таковой. На пересечении этих трёх достижений и родится ИскИн.

Однако всё может произойти и совсем не так…

На этой жизнеутверждающей ноте мы заканчиваем первую главу. Дальше мы рассмотрим три столпа искусственного интеллекта и современное состояние всех технологий, которые сегодня объединяются термином «искусственный интеллект». Речь, конечно, о слабом его варианте. Так что начнём…

# Глава 2

# ТРИ СТОЛПА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Теперь давайте погрузимся в самые глубины методов и технологий искусственного интеллекта. Само собой разумеется, что далее на протяжении этой и следующей главы мы будем рассматривать только искусственный интеллект в его слабом варианте. Но именно это позволит нам немного раскрыть своё сознание и подготовиться к четвёртой главе, в которой мы рассмотрим разные интересные аспекты исследования разума и сознания, в том числе и при помощи методов искусственного интеллекта. Ведь надо помнить, что одной из первичных задач рассматриваемого нами междисциплинарного направления исследований является моделирование как отдельных когнитивных функций, так и всего разума, для того чтобы изучать его на модели. Это и есть главная задача – получить модель, чтобы познать природу человеческого разума.

Но пока давайте рассмотрим операционное определение Искусственного Интеллекта, которое позволит точно определять, когда об этом предмете говорят что-то стоящее, а когда пытаются развесить на уши лапшу в жёлтой прессе. Итак, искусственный интеллект – это междисциплинарная область исследований и набор технологий, позволяющий создавать технические системы, решающие задачи, ранее доступные только человеку. Ну вот как-то так.

И давайте теперь определим термин «искусственная интеллектуальная система», или проще – «ИИ-система». ИИ-система – это техническая система, которая при помощи методов и технологий искусственного интеллекта решает задачу или набор задач, ранее доступных только человеку. При этом ИИ-системы обладают двумя важными качествами, которые будут дифференцировать их в ряду других подобных систем. Это автономность и адаптивность, так называемые «2А».

- Автономность – это свойство системы, позволяющее ей действовать и, главное, принимать решения самостоятельно, без дополнительных управляющих воздействий извне. Конечно, автономная система может принимать и исполнять такие внешние воздействия, но в условиях их отсутствия

она вполне может действовать и справляться со своими задачами самостоятельно.

- Адаптивность – это свойство системы, которое позволяет ей действовать в условиях изменчивости внешних воздействий. Если рассматривается техническая система, то адаптивность подразумевает возможность со стороны системы обрабатывать такие входные данные, которые не предусмотрены её изначальным проектом. Например, для нейросетевых систем это свойство обозначает, что система будет адекватно реагировать на такие стимулы, каких не было в обучающей выборке.

Это два главных свойства ИИ-систем, но они не единственные. Давайте изучим ещё три, и в итоге эти пять свойств будут очень точно описывать каждую ИИ-систему, так что их можно будет классифицировать и, соответственно, делать правильные умозаключения.

Третье свойство – то, в какой *реальности* функционирует ИИ-система. Тут имеется два варианта: это либо наша объективная реальность, в которой мы с вами живём, и тогда такая ИИ-система называется «роботом», либо это виртуальная реальность, которая сформирована для ИИ-системы программными методами. Виртуальная реальность может быть связана с нашей, а может быть и не связана. Виртуальный помощник функционирует в виртуальной реальности и воспринимает в ней только голосовые команды. Робот-пылесос работает в нашей реальности и воспринимает её многочисленными сенсорами, а также может воздействовать на неё своими исполнительными устройствами.

Четвёртое свойство – *общность* интеллекта. Здесь тоже имеется дихотомия: искусственный интеллект может быть узким и общим. Узкая ИИ-система решает только одну задачу или один класс задач. Она решает только ту задачу, для которой предназначена. Если ИИ-система играет в шашки, то играть в шахматы она не может, тем более не может водить машину. А вот персональный помощник умеет решать многие задачи, и это первый шаг к построению ИИ-системы общего плана, хотя в этом направлении человечество находится только в самом начале пути.

Наконец, пятое свойство. Это тоже дихотомия, которую мы уже рассмотрели в предыдущей главе, – слабый и сильный искусственный интеллект. Напомню для удобства: слабой называется ИИ-система, которая просто решает поставленную задачу, не задумываясь, не рефлексируя, не осознавая того, что делает. Слабый ис-

кусственный интеллект – это просто тот самый набор технологий, который позволяет решать нам задачи, относящиеся к прерогативе человека. Все известные нам на сегодняшний день ИИ-системы являются слабыми. А вот сильный искусственный интеллект обладает самосознанием, он осознаёт себя отдельной разумной сущностью. В этом и состоит его отличие. Проблема в том, что мы пока не обладаем необходимыми инструментами, которые позволяют нам ответить на вопрос, обладает ли ИИ-система самосознанием. Мы не можем этого сделать даже относительно другого человека.

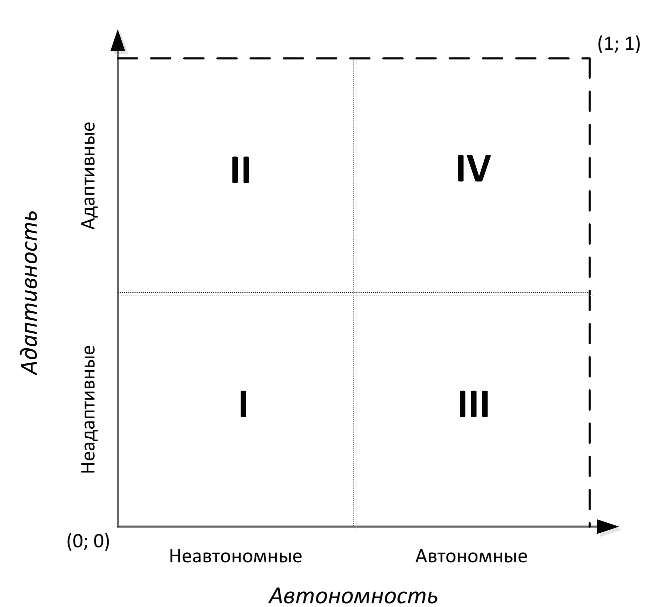
И вот теперь имеется пятимерное пространство, разделённое на тридцать два сектора, т. е. по одному сектору на каждый вариант пяти свойств, каждое из которых может принимать два значения. С одной стороны находится неавтономная неадаптивная узкая слабая ИИ-система, работающая в виртуальном мире. С другой стороны – автономный адаптивный общий сильный ИскИн, живущий рядом с нами. Человек, кстати, тоже находится в одном секторе с ИскИном, если только не принимать во внимание естественность его интеллекта. Ну, кстати, шкала «Искусственность – Естественность» является ещё одним измерением, которое, однако, в рамках темы этой книги никакого особого интереса не представляет.

Итак, повторим для закрепления ещё раз. Под интеллектуальностью искусственной (технической) системы понимается наличие у такой системы двух важных свойств. Во-первых, это возможность адаптации к изменяющимся условиям внешней среды при эксплуатации или изменяющимся условиям самой эксплуатации. Во-вторых, это высокая степень автономности её работы, в том числе и в части функциональности по принятию самостоятельных решений. Чем выше степень автономности и адаптивности искусственной системы, тем выше её интеллектуальность.

С другой стороны, под *интеллектуализацией* понимается процесс повышения степени интеллектуальности технической системы. Исходя из того что интеллектуальность является составной характеристикой из двух более простых свойств, интеллектуализация представляет собой траекторию развития системы в рамках своего жизненного цикла от низкого уровня к высокому по двухмерному пространству состояний.

Если рассмотреть относительные шкалы для описания и классификации технических систем по степеням их адаптивности и автономности, то можно представить непрерывное двумер-

ное пространство, в котором каждая точка может соответствовать определённой технической системе с заданными степенями адаптивности и автономности от точки (0, 0) – полностью неадаптивная и неавтономная система – до точки (1, 1) – совершенно адаптивная и автономная система. Эта мысль проиллюстрирована следующей схемой.

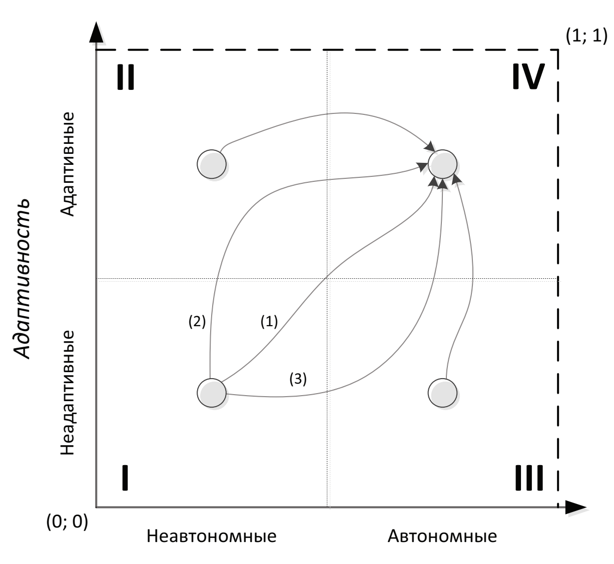



В качестве типовых примеров систем, которые находятся в квадрантах этого двумерного классификатора интеллектуальности систем, можно привести следующие.

1. *Робот-манипулятор на промышленном производстве*: обычно такой робот преднастроен на выполнение заданной последовательности действий (низкая адаптивность) и в принципе не предназначен для продуцирования каких-либо решений (отсутствие автономности).
2. *Робот-пылесос*: этот робот при попадании в незнакомое окружение обследует новое пространство и строит его модель для дальнейшего использования в своей работе (высокая адаптивность), но при этом выполняет только одну функцию уборки помещения с запуском по команде владельца или по триггеру в календаре (низкая автономность).

3. *Система поддержки принятия решений в заданной проблемной области*: такая система обычно уже преднастроена под определённые ситуации, и добавление новой ситуации требует существенной доработки системы (низкая адаптивность), при этом система сама готовит решения и объясняет их (высокая автономность).
4. *Персональный помощник*: система для планирования, напоминания, выполнения рутинных действий, которая постоянно обучается во взаимодействии со своим пользователем (высокая адаптивность) и при этом может самостоятельно запускать различные сценарии общения со своим пользователем и в рамках такого общения самостоятельно делать выбор в пользу того или иного варианта решения на основе прошлых предпочтений пользователя (высокая автономность).

Таким образом, интеллектуализация представляет собой процесс перевода (эволюции) технической системы из её текущего положения в пространстве «Адаптивность – Автономность» в четвёртый квадрант как можно ближе к точке (1; 1). Из квадрантов II и III такой переход может быть осуществлён непосредственно, в то время как в квадранте I может существовать три возможные траектории интеллектуализации системы, что показано на следующей схеме.



Таким образом, предполагается, что процесс интеллектуализации может быть применён к произвольной технической системе. И это довольно важный вывод – при помощи методов и технологий искусственного интеллекта возможно повышение интеллектуальности заданной технической системы посредством изменения её функциональности и внесения в неё двух важных свойств. Ну а делается это при помощи фактически трёх методов, которые рассматриваются далее в этой главе.

## Раздел 2.1. Символьные вычисления

Начнём с символьных вычислений. Это, если можно так выразиться, некоторая надстройка над методами представления знаний, которая позволяет представленными знаниями оперировать. Другими словами, символьные вычисления используются для вывода на имеющихся знаниях новых фактов в рамках некоторой проблемной области. Это кардинально отличается от нейросетевого подхода, который будет изучен далее, поскольку каждый шаг символьных вычислений можно объяснить, а сами методы основаны на формальной логике.

Первоначально символьные вычисления основывались на математическом понятии «формальная система», в рамках которого определялся алфавит для построения выражений (он является либо конечным, либо счётным), множество правильно построенных формул (подмножество выражений), множество аксиом (подмножество правильно построенных формул) и множество правил вывода (конечное множество отношений между правильно построенными формулами). Если не вдаваться в подробности, то в рамках формальной системы осуществляется полное абстрагирование от смысла слов естественного языка, на котором выражается какая-то теория, и её перевод в строгие формальные рамки с возможностью при помощи правил вывода синтаксического получения из аксиом и формул других формул. Кроме того, для формальных систем различных типов имеются дополнительные свойства, позволяющие эффективно определять выводимость формул, их разрешимость, непротиворечивость и полноту самой формальной системы.

Обычно множество правил вывода какой-либо формальной системы представляет собой набор посылок с заключениями. Другими

словами, каждое правило вывода – это формальное синтаксическое преобразование одной формулы в другую. Такое преобразование может быть записано при помощи нотации «ЕСЛИ… ТО…», а это, в свою очередь, обозначает, что правила вывода формальной системы могут быть описаны продукциями в их самом простом варианте без необходимости использования контекста, условий применимости и других более тонких свойств продукционной модели представления знаний.

От продукций, при помощи которых представляются знания, правила вывода формальной системы ещё отличаются и тем, что выражения в посылках и заключениях таких правил рассматриваются в качестве синтаксических конструкций, а сам вывод представляет собой манипулирование этими конструкциями. Такой подход позволяет реализовать универсальную машину вывода, а процесс интерпретации смысла или семантику вынести на более высокий уровень в рамках интеллектуальной системы. Другими словами, результаты осуществлённого вывода воспринимаются специальным интерпретирующим модулем системы, который посылает команды исполнительным устройствам. Последние уже взаимодействуют с окружением интеллектуальной системы. И для того чтобы она могла называться интеллектуальной, такое взаимодействие должно соответствовать разумному ожиданию акторов, с которыми система взаимодействует.

Необходимо отметить, что некоторые формальные системы содержат счётные бесконечные множества аксиом, правильно построенных формул и даже правил вывода. Это значит, что в явном виде их перечислить невозможно. Вместо явного перечисления используются так называемые схемы – аксиоматические схемы и схемы правил вывода. Так что универсальная машина вывода должна иметь возможность работать и со схемами, на основании которых можно создавать счётные бесконечные множества правил вывода и, соответственно, правильно построенных формул.

Итак, ранее уже были упомянуты логические правила вывода, на которых основывается работа универсальной машины вывода. Во-первых, это правило *ModusPonens*, которое звучит как «Если есть правило, что из некоторого факта А следует заключение Б, и если при этом факт А истинен, то можно сделать вывод, что заключение Б тоже истинно». Это правило вывода предназначено для осуществления прямого вывода, когда есть набор фактов (в случае формальной системы – аксиом) и из них необходимо

вывести максимальное количество истинных заключений (в формальных системах – правильно построенных формул).

Во-вторых, есть правило *ModusTollens*, которое звучит как «Если есть правило, что из некоторого факта А следует заключение Б, и при этом установлено, что заключение Б ложно, то можно сделать вывод, что факт А тоже ложен». Это правило вывода используется в обратной стратегии вывода, когда наблюдается некоторое наличное состояние объекта наблюдения в рамках проблемной области и необходимо понять, что могло бы привести к этому состоянию, какие причины лежат в его основе. Обратный вывод обычно даёт вероятностные оценки возможных причин.

Описанные правила вывода основаны на обычной аристотелевой логике. Эти правила использовались в рассуждениях в виде отдельных типов силлогизмов ещё в античном мире. Реализация машины вывода, которая соблюдает такие правила, является делом достаточно простым. Тем не менее это очень мощный формализм, который позволяет решать многие задачи. Однако в конце XX века были предложены новые формализмы, которые расширяют аристотелеву логику и позволяют ещё более тонко подходить ко многим проблемным областям. Рассмотрим некоторые из них.

Первый важный формализм – нечёткая логика и связанные с ней понятия «лингвистическая переменная» и «нечёткая переменная». По мере осуществления попыток формализовать при помощи аристотелевой логики знания о более или менее сложной проблемной области становилось понятно, что простой «чёрно-белый» вариант логики с двумя значениями истинности не может описать всю гамму возможностей при рассмотрении человеком вариантов решения задач. Дело в том, что человек обычно решает задачи не в идеальном мире, а в условиях неполноты информации, нечёткости определений, неточности измерений и т. д. И все такие НЕ-факторы практически невозможно описать при помощи аристотелевой логики. На помощь пришли многозначные логики и, как их апофеоз, бесконечнозначная нечёткая логика. Последняя, к примеру, позволяет осуществлять логический вывод при наличии фактов, не совпадающих с посылками продукций, но машина вывода с возможностью обрабатывать нечёткость всё так же будет способна получить результат, чаще всего вполне приемлемый.

Нечёткий вывод тоже основан на правилах ModusPonens и ModusTollens, но они несколько модифицированы. Первое вы-

глядит как: «Если есть правило, что из некоторого факта А следует заключение Б, и при этом имеется некоторый факт А*, не совсем совпадающий с фактом А, то можно сделать вывод, что заключение Б должно принять вид Б*». В качестве примера можно рассмотреть такой вывод. Пусть есть правило «Если температура окружающей среды низкая, то длительность прогулки короткая». На вход нечёткой машине вывода подаётся факт: «Температура окружающей среды очень низкая», – и при помощи имеющегося правила и этого факта машина вывода делает вывод: «Длительность прогулки очень короткая». Это довольно тривиальный пример, который тем не менее показывает, что нечёткий вывод при помощи небольшого количества правил охватывает огромное число возможностей, в том числе и таких, которые вообще не были ранее предусмотрены разработчиком базы знаний с продукциями. Нечёткий вывод нашёл широкое применение в системах автоматического управления, но он также применяется и в интеллектуальных системах иных классов.

Другим важным классом формальных логик являются так называемые модальные логики. Это довольно обширный класс, и даже простейшая аристотелева логика может быть представлена в виде модальной. Но формализм модальных логик хорош тем, что при помощи него можно обобщать различные понятия и отношения объективной реальности. В частности, такая частная логика, как темпоральная, позволяет формально рассуждать о вопросах, связанных со временем. В этой логике высказывания также имеют временную привязку – некоторые могут происходить одновременно, другие следовать друг за другом. Другой важной модальной логикой является пространственная логика, которая так же, как и темпоральная, имеет средства выражения пространственных отношений между объектами и высказываниями. Такие высказывания, как «Красный куб находится ближе синего цилиндра», просто формализуются при помощи пространственной логики и очень сложно при помощи аристотелевой. Существуют логики и иных модальностей, но не каждая из них общеприменима, а некоторые эффективны только для каких-либо узких проблемных областей.

Описанные модальные логики имеют важнейшее практическое применение в рамках так называемого пространственно-временного вывода. Чаще всего этот вариант машинного вывода используется в робототехнике, поскольку именно роботы (не только

человекоподобные автономные, но и производственные манипуляторы) должны ориентироваться в пространстве и времени. Но также пространственно-временной вывод используется в сложных проблемных областях, в которых требуются рассуждения о том, что было или будет. Чаще всего такие проблемные области относятся к динамическим системам различной природы. В качестве примера можно привести медицину – динамика развития заболевания должна рассматриваться через призму временных отношений, а потому использование временной модальной логики неизбежно.

Важнейшим свойством символьных вычислений является возможность объяснить вывод, полученный при помощи них. Если есть начальные факты, есть последовательность рассуждений и есть множество правил вывода, то можно объяснить, почему из начальных фактов были получены такие-то результаты. Это очень важное свойство интеллектуальных систем, и оно в большей мере соответствует нисходящему подходу в построении искусственного интеллекта.

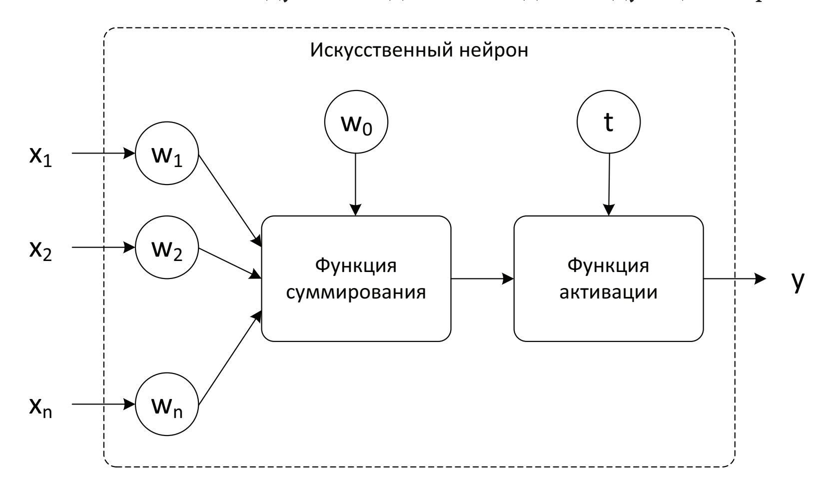
## Раздел 2.2. Искусственные нейронные сети

Искусственная нейронная сеть – это в первую очередь математическая модель машинного обучения, которая решает задачи примерно по тем же принципам, что и биологические нейронные сети, состоящие из нервных клеток. Искусственные нейронные сети имеют программную или аппаратную реализацию, поэтому чисто абстрактная математическая модель воплощена в действительности в виде вычислительной системы. В то же самое время искусственные нейронные сети представляют собой самую широко распространённую модель машинного обучения и одновременно с этим являются типичным подходом в рамках восходящего искусственного интеллекта.

Искусственная нейросеть состоит из множества взаимодействующих друг с другом искусственных нейронов, чаще всего собранных в слои так, что нейроны одного слоя получают информацию только от нейронов предыдущего слоя и передают информацию на следующий слой. Сам по себе искусственный нейрон представляет собой *очень* упрощённую модель биологического нейрона. У искусственного нейрона есть множество входов, один выход,

а также две функции: суммирования и активации. Впрочем, существуют и другие типы искусственных нейронов, в том числе и таких, работа которых зависит от времени. Фактически каждый искусственный нейрон решает задачу простой логистической регрессии (в случае если его функция активации представляет собой сигмоиду).

Первую модель искусственного нейрона предложили Мак-Каллок и Питтс в 1943 году. Эта модель выглядит следующим образом.




Чаще всего в качестве упрощения считается, что входные переменные $x_1$, $x_2$, ... , $x_n$ и выходное значение $y$ принимают значения из интервала [0, 1]. Функция суммирования представляет собой обычную взвешенную сумму, результат которой равен сумме произведений каждого входного значения $x_i$ на свой вес $w_i$, а некоторый специальный вес $w_0$ добавляется всегда. Наибольший интерес представляет функция активации, которая может принимать различные формы, так что свойства как искусственного нейрона, так и всей нейросети часто зависят от выбора формы функции активации.

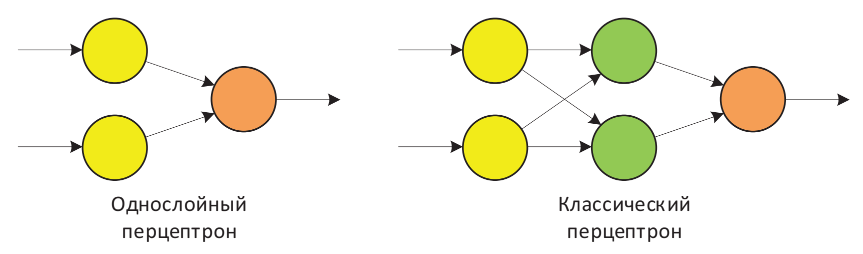
Результат функции суммирования передаётся на вход функции активации. В свою очередь, её результат является выходом нейрона $y$, который передаётся на вход следующим нейронам. Обычно, как уже было указано, функцию активации подбирают так, чтобы область её значений лежала в интервале [0, 1], хотя это и необязательно. В первых реализациях искусственных нейронных сетей

функция активации вообще была дискретной с областью определения {0, 1}, и такие нейросети неплохо справлялись с задачами классификации. Сегодня в сетях глубокого обучения могут использоваться разные функции активации у разных нейронов, но обычно все они входят в класс сигмоид – это и логистическая функция, и гиперболический тангенс, и функция Гаусса, и многие другие. Выбор функции активации зависит от многих параметров задачи, и часто архитектор, проектирующий искусственную нейросеть, должен проявить для этого немалый уровень творчества.

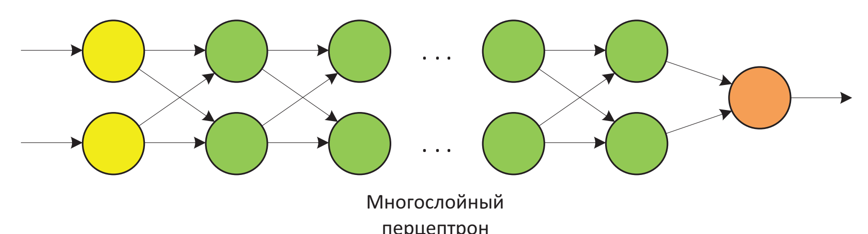
Со времён, когда первые разработчики искусственных нейронных сетей сделали свои первые искусственные нейросети, были созданы многочисленные модификации модели, которые подходят для решения более специфичных задач. Конечно, для погружения в теорию искусственных нейросетей необходимо читать специализированную литературу и изучать курсы, но здесь можно кратко охарактеризовать многие интересные модели и решаемые ими задачи. Так что начнём…

Несомненно, первой реализованной моделью искусственной нейронной сети был перцептрон Ф. Розенблатта, который тот предложил в 1957 году. Эта модель была реализована «в железе», первым в мире «нейрокомпьютером» стал компьютер Марк-1, построенный под руководством Розенблатта в 1960 году. Перцептрон – это простая нейронная сеть с тремя слоями: входным, скрытым и выходным. Таким образом, перцептрон реализует простейшую кибернетическую машину с сенсорами (слой входных нейронов), управляющим устройством (слой скрытых нейронов) и аффекторами (слой выходных нейронов). В перцептроне используется пороговая передаточная функция и прямое распространение сигнала. Как математическая модель перцептрон уже был достаточно мощным формализмом для решения большого количества задач, поскольку он на основе обучения позволял классифицировать, кластеризовать и прогнозировать, т. е. решать большинство классических задач машинного обучения. Впрочем, после публикации книги М. Минского и С. Паперта «Перцептроны», в которой авторы показали принципиальную невозможность для перцептрона решить некоторые задачи (сюрприз – задача «XOR», традиционно включаемая в класс нерешаемых перцептроном, на самом деле к таковым не относится), постепенно интерес к перцептрону снизился, и большее внимание стала получать нисходящая парадигма в искусственном интеллекте, при этом сам Марвин Минский был

её оппонентом. Тем не менее сегодня с развитием математического аппарата и средств вычислительной техники интерес к перцептрону и его расширениям вновь вырос.




Необходимо отметить, что в процессе развития перцептронов появились некоторые расширения первоначальной модели, предложенной Ф. Розенблаттом. Самый простой классификатор основан на подсчёте слоёв в перцептроне: однослойный, с одним скрытым слоем (классический) и многослойный. Все эти типы были в своё время описаны Розенблаттом. Другая авторская классификация включала: элементарный перцептрон, простой перцептрон, перцептрон с последовательными связями, перцептрон с перекрёстными связями, перцептрон с обратными связями, перцептрон с переменными связями. Первые три класса были описаны самим автором, а следующие три развиты в дальнейшем при детальной проработке модели искусственных нейронных сетей.




Из-за первоначальной неразберихи в терминологии и повышенных ожиданий, которые появились в отношении перцептрона и модели искусственного нейрона, Дэвидом Румельхартом был предложен новый класс перцептронов, которые сейчас называются «многослойными перцептронами Румельхарта» и отличаются от многослойных перцептронов Розенблатта тем, что для обучения в них используется метод обратного распространения ошибки, в то время как у Розенблатта использовался метод коррекции

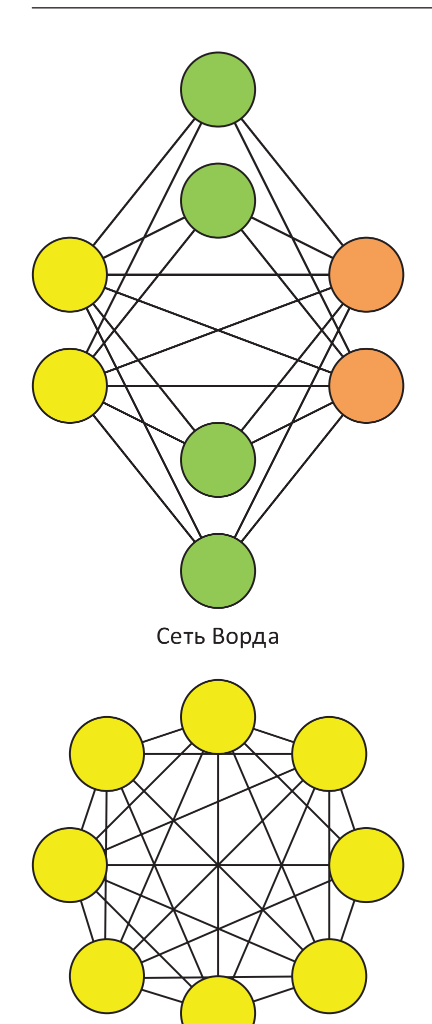
ошибки. Есть ещё несколько отличий, в частности у Румельхарта в качестве функции активации используется сигмоида, а число обучаемых слоёв больше одного.

Фактически перцептрон – это самый простой пример нейронной сети прямого распространения. Другим интересным примером является сеть ELM, экстремальная обучающаяся машина. В этой нейросети нейроны не располагаются в слои, а связаны друг с другом случайным образом. Выделяются только входной и выходной слои, а остальные нейроны находятся между ними и связаны друг с другом именно случайно. Обучение сети также производится методом обратного распространения ошибки. А вот если обучение производится при помощи обновления состояния нейронов по результатам наблюдения за их работой, за порядком активации, то это уже нейронная эхо-сеть, ESN.

Далее некоторое время модель искусственных нейронных сетей развивалась в сторону тонкого подбора функций активации (например, в сплайн-модели Хакимова функция активации реализована в виде сигмоидального сплайна, либо в сети Брумхеда–Лоу используются радиально-базисные функции), типов весовых коэффициентов между нейронами и других параметров, не влияющих на структуру сетей. Это не позволяло выходить на новые объёмы данных или решать новые задачи, но было обусловлено скорее тем, что на существующих в те времена вычислительных мощностях было затруднительно реализовывать нейронные сети больших размеров.

Однако перцептрон получал развитие и с точки зрения изменения структуры. Так, например, Джефф Элман и Майкл Джордан предложили свои варианты изменения перцептрона и включения в него обратных связей так, чтобы получилась рекуррентная сеть. В варианте Джордана на вход многослойного перцептрона также подаются выходные значения с задержкой от одного и более тактов. В нейронной сети Элмана используется такой же подход, только рекуррентные связи с задержкой в несколько тактов подаются не с выходных нейронов, а со скрытых слоёв многослойного перцептрона. Оба варианта структурного развития перцептрона обладают определённой памятью. Впрочем, различными вариантами памяти обладают все рекуррентные сети.

Ещё один интересный вариант нейросети с небольшим количеством скрытых слоёв – сеть Ворда, в которой имеется только один внутренний слой, но его нейроны разбиты на группы. В каждой


Сеть Ворда

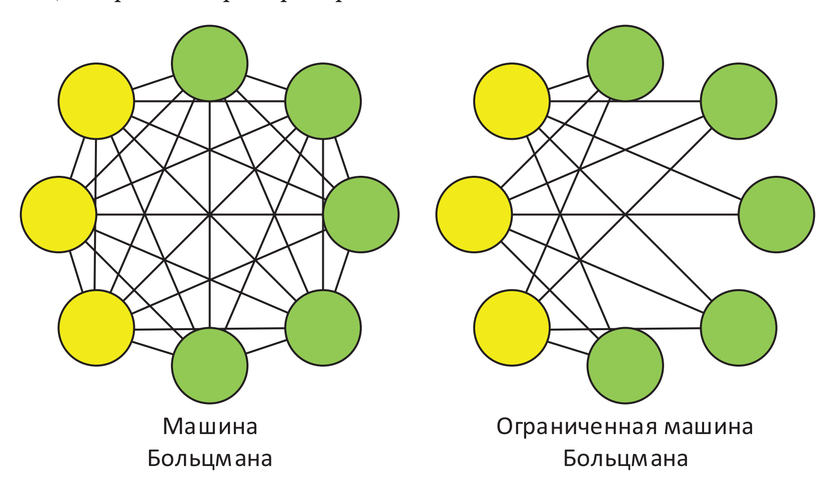
Сеть Хопфилда

группе используется своя передаточная функция, и результаты работы каждого блока передаются на выходной слой, который как бы рассматривает вход с разных точек зрения. Кроме того, входной слой может напрямую замыкаться на выходной, и это тоже позволяет добавить гибкости сети. Топология конкретной реализации сети Ворда определяется количеством блоков в скрытом слое и наличием прямого замыкания входных нейронов на выходные. Сеть показывает отличные результаты для задачи распознавания образов.

В 1982 году Джон Хопфилд предложил интересный вариант однослойной нейросети, который позволял достаточно просто решать некоторые задачи оптимизации или предоставлять механизм автоассоциативной памяти. В этой сети имеются только входные нейроны, каждый из которых соединён со всеми другими (т. е. сеть представляет собой полносвязный граф). Обучение сети также своеобразно – для всех связей между нейронами необходимо аналитически рассчитать веса, и это значит, что обучение длится один цикл. Обучение (или аналитический расчёт весов) как бы сохраняет в весовых коэффициентах сети набор эталонных образцов, которые сеть должна запомнить. После того как сеть обучена, её можно запускать в рабочем режиме, и он также отличается от традиционного единичного прогона. Сеть Хопфилда работает до тех пор, пока её следующее состояние не будет равно

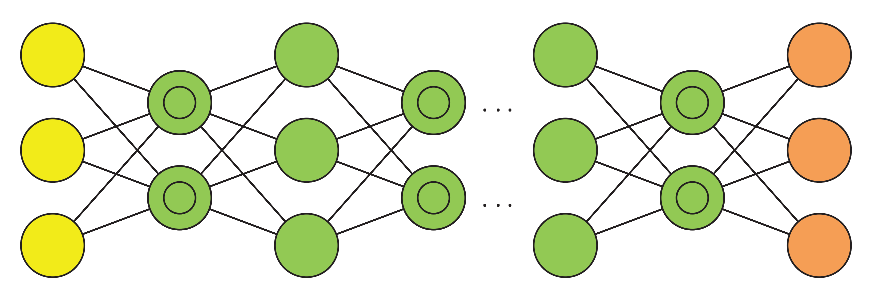
предыдущему, т. е. сеть должна достигнуть равновесия. Достигнутое равновесное состояние обязательно соответствует одному из запомненных эталонных образцов. Тем самым осуществляется распознавание и даже восстановление повреждённых образцов.

Развитием сети Хопфилда стала так называемая неограниченная машина Больцмана, которая представляет собой стохастический вариант этой сети. Машина Больцмана представляет собой рекуррентную нейронную сеть, которая для обучения использует алгоритм имитации отжига. Неограниченная машина позволяет решать сложные комбинаторные задачи, однако её практическое применение затруднительно или даже невозможно из-за комбинаторного взрыва в части гигантского количества связей между нейронами. Однако если использовать ограничение на количество связей между нейронами, то машина Больцмана становится менее мощной, но её использование вполне возможно. В частности, из каскадов из машин Больцмана составляются глубокие сети доверия – сначала сеть обучается при помощи стандартного алгоритма для машин Больцмана, а потом дообучается при помощи обратного распространения ошибки.




Итак, рассмотрим каскадное объединение ограниченных машин Больцмана, при котором скрытые нейроны одной машины становятся входными для следующей. Именно так формируется глубокая сеть доверия. Такая сеть также обучается без учителя, причём обучение происходит послойно. Как и в случае сети Хоп-

филда и машины Больцмана, сеть глубокого доверия осуществляет кластеризацию и восстановление входов, а при дополнительном обучении с учителем осуществляется и классификация, которая может быть очень точной. Именно на глубоких сетях доверия были построены первые алгоритмы глубинного обучения.



Глубокая сеть доверия

Ещё один пример глубокой нейронной сети прямого распространения – это глубокая остаточная сеть (DRN). В ней количество скрытых слоёв может быть очень велико, и прямые связи от слоя передаются не только на следующий слой, но и на дальнейшие слои. При этом на каждый слой также передаётся входная информация в неизменённом виде, так что каждый слой получает как бы пару (входная информация, выходная информация предыдущего слоя). Такая архитектура может очень неплохо распознавать образы, однако было показано, что подобная нейросеть тождественна рекуррентным нейросетям без учёта времени.

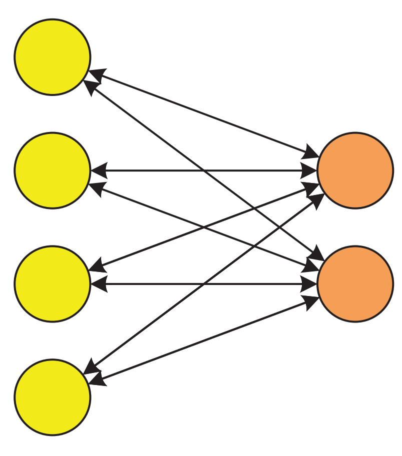
Так что теперь более подробно рассмотрим эти самые рекуррентные нейронные сети, т. е. такие, в которых возможно либо замыкание выхода нейрона на его же вход, либо создание обратных связей от нейронов дальних слоёв к нейронам ближних. При этом значение на рекуррентные связи подаётся с определённой временной задержкой, измеряемой в тактах работы нейронной сети. В принципе, и сеть Хопфилда, и машина Больцмана уже сами по себе являются рекуррентными сетями, но они представляют самый примитивный их класс. Теперь пришло время познакомиться с более продвинутыми.

Одна из самых первых рекуррентных нейронных сетей была предложена в 1987 г. – так называемая сеть Хэмминга, которая является расширением сети Хопфилда и решает задачу классификации бинарных векторов, критерием различия которых является

расстояние Хэмминга. Сеть Хэмминга представляет собой трёхслойную сеть, при этом количество нейронов во втором и третьем слоях равно количеству классов классификации, и со второго и третьего слоёв есть рекуррентные связи – со второго на первый, а с третьего на третий же. После обучения сеть выдаёт номер класса, которому принадлежит образец, поданный на вход. Эта сеть очень эффективно работает на довольно узком классе задач. В частности, она может использоваться для распознавания чёрно-белых изображений: штрих-кодов, цифр индекса и др.

В 1988 г. Барт Коско разработал тип нейронной сети, названный им «двунаправленной ассоциативной памятью» (а сегодня она называется сетью Коско). Это вариант сети Хопфилда, в котором входными является часть нейронов, а выходными – оставшиеся нейроны. Все нейроны входного и выходного слоёв соединены друг с другом двунаправленными связями, которые обучаются так же, как и для сети Хопфилда. Сеть Коско работает на тех же задачах, что и сеть Хопфилда, но делает это немногим более эффективно. Также разработаны некоторые подтипы этой сети, в частности адаптивный вариант, в котором осуществляется изменение весов связей в процессе работы. Тем самым адаптивная сеть Коско позволяет идентифицировать искажённые и зашумлённые образцы.

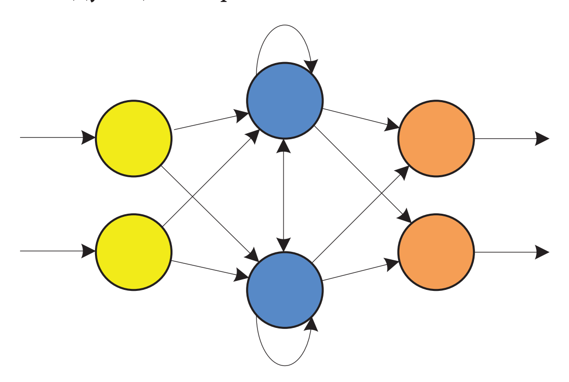
Вообще, рекуррентными нейронными сетями называют любые искусственные нейросети, в которых есть обратные связи от нейронов более глубоких слоёв к нейронам менее глубоких, либо замыкание нейронов самих на себя или на нейроны того же самого слоя. При этом такая архитектура связей между нейронами и слоями может создаваться и в уже рассмотренных типах сетей, в которых изначально рекуррентных связей не предполагалось. Подобные сети могут запоминать информацию, организовывать ци-



Сеть Коско

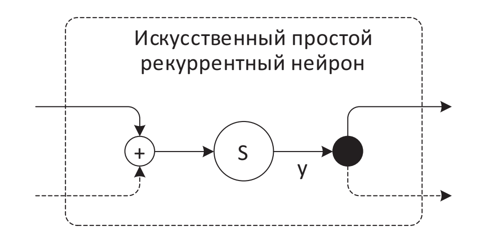
клы и иметь другие интересные свойства. Однако из-за сложности анализа поведения таких моделей большая часть предложенных или потенциальных архитектур изучена плохо.

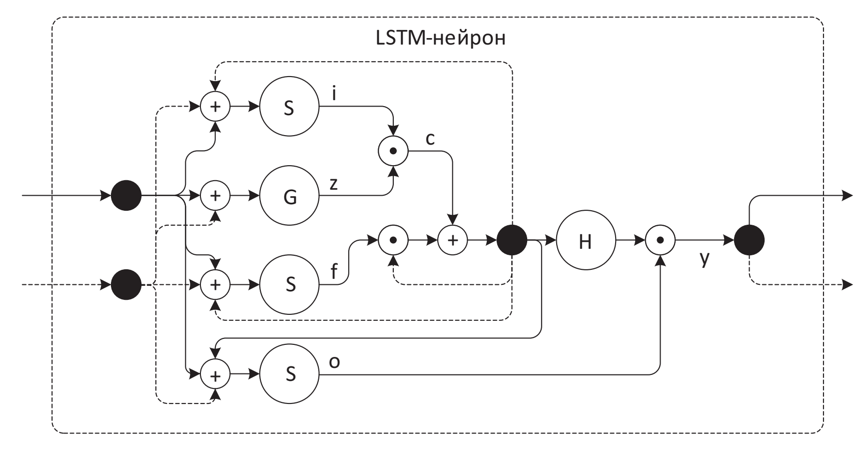
Общий вид архитектуры рекуррентной нейронной сети можно представить следующим образом.



В процессе исследований рекуррентных нейронных сетей стало ясно, что с обычными искусственными нейронами они работают не совсем эффективно – обычные нейроны с течением времени как будто бы «размываются», теряют своё обучение. Это не позволяет эффективно использовать обычные нейроны в рекуррентных нейронных сетях для задач классификации и, главное, прогнозирования временных рядов. И тогда появилась идея по разработке новых типов искусственных нейронов. Одним из наиболее знаменитых в этом плане является LSTM-нейрон (от *англ.* LongShort-TermMemory – «долгая краткосрочная память»).

Итак, Сепп Хохрайтер и Юрген Шмидхубер в 1997 г. предложили новый вариант искусственного нейрона, который затем был улучшен Феликсом Герсом в 2000 г. Искусственная нейросеть, составленная из LSTM-нейронов, является универсальной в смысле Тьюринга, т. е. при помощи неё можно вычислить любую вычислимую функцию (матрицу весов такой нейросети можно воспринимать как программу). Именно такие нейросети сегодня являются фундаментальной основой для построения решений ведущих компаний, работающих в области искусственного интеллекта. Сам же LSTM-нейрон может быть представлен при помощи следующей диаграммы.




Представленные выше диаграммы нарисованы с использованием следующих условных обозначений:

- обычная стрелка – входные и выходные связи нейрона, поток значений между функциями внутри нейрона;
- пунктирная стрелка – входные и выходные связи нейрона, поток значений между функциями внутри нейрона с задержкой по времени (на один или более тактов работы нейронной сети);
- плюс – функция суммирования входных значений;
- точка – функция перемножения входных значений;
- чёрный круг – дубликатор потока значений;
- гейт **S** – функция активации, обычно представляющая собой сигмоиду;
- гейт **G** – функция активации для входного потока, обычно являющаяся гиперболическим тангенсом;
- гейт **H** – функция активации для выходного потока, обычно тоже гиперболический тангенс;
- поток i – результат активации для входного потока;

- поток z – результат активации для рекуррентного входного потока (обычно с нейронов того же слоя);
- поток f – результат активации для гейта забывания;
- поток o – результат активации для выходного потока;
- поток y – результат работы LSTM-нейрона в целом.

На сегодняшний день, как уже сказано выше, LSTM-нейрон является самой продвинутой моделью биологического нейрона. Однако он всё так же имеет одно важное упущение – нейроны в нервной системе млекопитающих получают информацию не только при помощи передачи электрических импульсов от дендритов через аксоны к следующим нейронам, но ещё используются различные нейромедиаторы в синаптических щелях, а также гуморальное воздействие на весь нейрон в целом. Кроме того, до сих пор не до конца изучена роль глиальных клеток, которых в нервной системе на порядок больше, чем нейронов. Другая сложность, которая проявляется при реализации искусственных нейросетей на LSTM-нейронах, – их относительно высокие требования к вычислительным ресурсам.

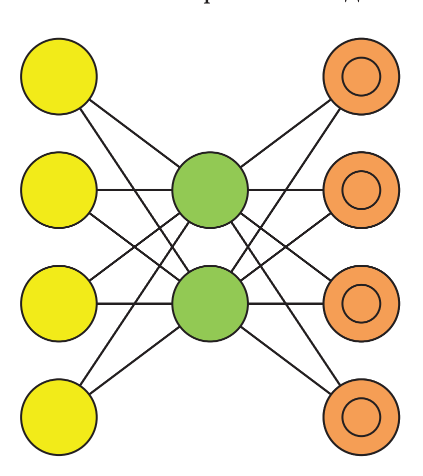
Из-за описанных сложностей с LSTM-нейроном был разработан так называемый GRU-нейрон (от *англ.* gatedrecurrentunit – «рекуррентный нейрон с гейтами»), который практически повторяет структуру LSTM-нейрона, но в нём нет выходного гейта **H** и, соответственно, всех используемых этим гейтом потоков. Это позволило существенно снизить вычислительную сложность работы такого искусственного нейрона, однако он так же эффективен для отдельных задач, как и LSTM-нейрон (в частности, на GRU-нейронах хорошо решаются задачи синтеза – моделирование музыки или моделирование текста и разговора).

Мы изучили сети прямого распространения и рекуррентные сети, и уже по пройденному материалу видно, как широко это направление исследований. Так что теперь мы можем перейти к более прикладным вариантам, на которых сегодня решается большинство задач, где используются искусственные нейронные сети.

Самый первый интересный вариант – так называемый «автокодировщик». Этим словом обозначается специальная архитектура искусственной нейронной сети с одним скрытым слоем без обратных связей. Главная особенность такой сети заключается в том, что количество нейронов на входном и выходном слоях одинаково, а на скрытом слое – меньше. Это ограничение необходимо для того, чтобы сеть не зафиксировалась на тривиаль-

ном варианте в процессе обучения без учителя, которое для автокодировщиков заключается в получении на выходном слое результатов, как можно более похожих на входные параметры сети. Другими словами, обучение нейронов скрытого слоя в таких условиях позволяет автокодировщику как бы осуществлять обобщение, сжимающее кодирование входной информации, отсюда и такое наименование.

Автокодировщики часто используются для последовательного обучения слоёв искусственных нейронных сетей. При таком подходе каждый новый слой представляется в качестве скрытого слоя автокодировщика, при этом входным слоем является предыдущий слой, обученный на предыдущем же шаге. Размерность каждого следующего слоя снижается, и тем самым нейросеть должна находить всё больше и больше обобщений, она должна всё сильнее и сильнее повышать степень абстракции кодируемой информации. После такого обучения нейросеть на каждом скрытом слое содержит нейроны, которые отвечают за очень глубокие уровни абстракции входной информации, и чем глубже слой, на котором находится нейрон, тем выше уровень абстракции. Сегодня это один из основных подходов в технологии глубокого обучения.


Автокодировщик

Перед переходом к рассмотрению глубоких нейронных сетей остаётся отметить несколько особенных вариантов автокодировщиков.

- *Вариационный автокодировщик* – автокодировщик, в рамках которого делаются сильные предположения относительно распределения скрытых переменных. Такие автокодировщики используют вариационное исчисление для обучения представлению скрытых переменных, что приводит к дополнительной потере компонентов при использовании специализированного алгоритма обучения. Он предполагает, что данные сгенерированы направленной графической

моделью, и автокодировщик при обучении строит аппроксимацию апостериорного распределения.

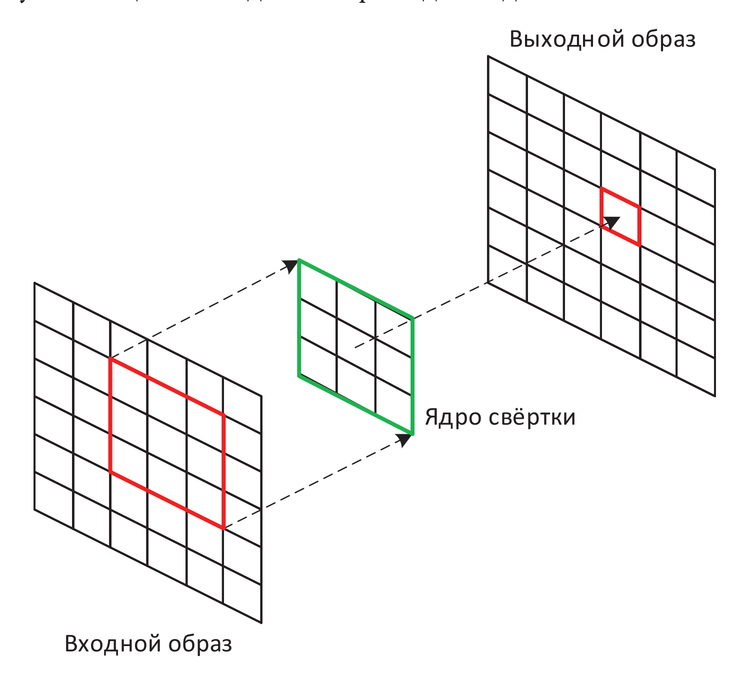
- *Автокодировщик, удаляющий шум* – автокодировщик, который принимает частично «повреждённые» входные данные, а обучение осуществляется для получения неискажённого выхода. Такие автокодировщики восстанавливают так называемое «хорошее представление», которое может быть получено из повреждённого выхода так, что достаточно точно воспроизводит ожидаемый выход.
- *Разрежённый автокодировщик* – автокодировщик, который во время обучения имеет большее число скрытых нейронов, чем входных. Это создаёт так называемую «разрежённость» в нейронах, которая накладывается и на выходные нейроны. Иногда это бывает полезно при решении задач классификации.
- *Контрастный автокодировщик* – автокодировщик, который явным образом добавляет регуляризацию своей целевой функции, что заставляет нейронную модель исследовать функцию, которая является устойчивой к небольшим изменениям входных значений.

Теперь, наконец-то, перейдём к рассмотрению искусственных нейронных сетей, которые получили название «свёрточных» (а также двойственных им – «развёрточных»). Такие нейросети получили в последнее время самое широкое распространение, поскольку являются базовым строительным блоком для построения систем «глубокого обучения». Рассмотрим несколько примеров.

Свёрточная сеть – ещё одна специальная архитектура искусственных нейронных сетей, которая очень эффективно решает задачу распознавания образов. Сама по себе архитектура была предложена в 1988 г. французским учёным Яном Лекуном, который взял некоторые идеи из устройства нервной системы млекопитающих и, в частности, зрительной коры головного мозга. Структура свёрточной нейронной сети состоит из большого количества слоёв двух чередующихся типов – свёрточных и субдискретизирующих. Обычно работа такой сети описывается как постепенный переход от конкретных особенностей распознаваемого образа к абстрактным деталям всё выше и выше по иерархии уровней абстракции. Например, в применении к изображениям это обозначает, что первые слои обнаруживают в частях изображения различные типовые элементы вроде разнонаправленных отрезков, а на последних слоях нейроны уже активируются в от-

вет на наличие в составе изображения каких-либо форм, конкретных объектов или образов.

Суть операции свёртки, из-за которой сеть получила своё название, заключается в том, что каждый фрагмент входного образа поэлементно умножается на матрицу свёртки, результат этого перемножения суммируется, и итоговая сумма передаётся в аналогичную позицию выходного образа для заданного слоя.




Каждый слой свёртки передаёт свой выход на слой субдискретизации, и это важная особенность рассматриваемой архитектуры. Субдискретизация нелинейно уплотняет карту признаков так, что группа пикселей (обычно 3×3) преобразуется в один пиксель, и это делается при помощи нелинейного преобразования (обычно используется функция максимизации). Суть этой операции в том, чтобы получить карту признаков с предыдущего слоя свёртки и «огрубить» изображение. Другими словами, происходит отказ от ненужных деталей, и нейронная сеть начинает рассматривать всё более и более абстрактные признаки исходного изображения. Кроме максимизации в качестве операции субдискретизации, могут использоваться и другие, и от этого зависят свойства получа-

емой нейронной сети, лучше или хуже подходящие для решения поставленной задачи.

Наибольший интерес вызывает то, что в операции свёртки матрица или ядро свёртки изначально не закладывается разработчиком нейронной сети, а самостоятельно подбирается в процессе обучения нейронной сети (чаще всего методом обратного распространения ошибки). В итоге для каждого свёрточного слоя получается большое количество карт признаков – отрезки и дуги, направленные под разными углами, границы между сплошными средами, точки и иные примитивы; а на более глубоких слоях – лица, животные, автомобили, здания и т. д. Количество карт признаков для каждого слоя является метапараметром сети и определяется разработчиком.

Для чего нужна свёрточная нейронная сеть? Как показали эксперименты, свёрточные нейросети:

- являются одним из лучших методов по распознаванию образов;
- по сравнению с полносвязными архитектурами (например, с перцептронами), требуют намного меньшего количества вычисляемых в процессе обучения параметров, из чего следует возможность построения сетей с большим количеством слоёв;
- позволяют эффективно распараллеливать свои вычисления;
- обладают устойчивостью к поворотам и сдвигам распознаваемых образов.

Если про искусственные нейронные сети можно написать отдельную книгу, то про свёрточные нейронные сети можно написать ещё одну, поэтому оставим их и перейдём к следующему типу архитектуры – развёрточным сетям.

Развёрточные нейронные сети – это как бы перевёрнутые задом наперёд свёрточные сети. Они были предложены Мэтью Зайлером для простой задачи – анализа работы свёрточных нейронных сетей. Дело в том, что глубинное обучение задействует воистину огромное количество слоёв, а в свёрточных нейросетях эти слои ещё и разных типов, и субдискретизация осуществляется по разным каналам, а с учётом того, что чем глубже и полносвязнее сеть, тем сложнее понять, как она обучается, необходим хотя бы какой-то эвристический механизм или инструмент для оценки того, правильно ли обучается свёрточная нейросеть и не переобучилась ли она.

Если говорить самым простым языком, то развёрточная нейронная сеть строит иерархические представления свёрточной

нейросети, к которой подключена. При этом в рассмотрение принимаются все карты признаков и параметры, которые были получены во время обучения свёрточной нейросети. Развёрточная нейронная сеть как бы пытается восстановить тот сигнал, который распознавала свёрточная нейросеть, но из-за большого количества нелинейных и необратимых преобразований это удаётся сделать лишь частично. Однако развёрточная нейросеть подключена к каждому свёрточному слою свёрточной нейросети и восстанавливает изображения для всех свёрточных слоёв, обучаясь параллельно. В итоге получается нейронная сеть, которая позволяет «видеть» то, как обучена свёрточная нейронная сеть и хотя бы отчасти объяснять результаты.

Перейдём к следующему важному варианту архитектуры искусственных нейронных сетей. Это генеративно-состязательная сеть (GAN), которая предназначена для обучения без учителя. Генеративно-состязательная нейронная сеть составлена из двух подсетей. Одна из них (сеть G – генеративная сеть) генерирует различные образцы, а вторая (сеть D – дискриминационная сеть) – отличает правильные образцы от неправильных. Для целей машинного обучения сети G и D имеют противоположные задачи, и между ними организуется «антагонистическая игра», т. е. игра, в которой участвуют два игрока, выигрыши которых противоположны. Сеть G генерирует образец (например, изображение), начиная со смешивания некоторых исходных образцов, используя случайные значения своих скрытых параметров (весовых коэффициентов). После генерации образца сеть D пытается отбраковать те образцы, которые выглядят неприемлемо, и результат работы этой сети подаётся на вход сети G, которая при помощи метода обратного распространения ошибки перестраивает свои весовые коэффициенты так, чтобы «обмануть» сеть D. В итоге на каждой итерации генерируемые образцы становятся всё лучше и лучше.

Обычно в качестве сети D берётся свёрточная сеть глубокого обучения – ну просто потому, что именно свёрточные сети, как уже описано, лучше всего занимаются распознаванием образов, а дискриминационная сеть должна именно распознавать образы, которые создаёт генеративная сеть. А в качестве последней используются те сети, которые могут генерировать образы. И вообще говоря, сегодня именно генеративные нейронные сети находятся на пике интереса, и здесь могут быть самые разные

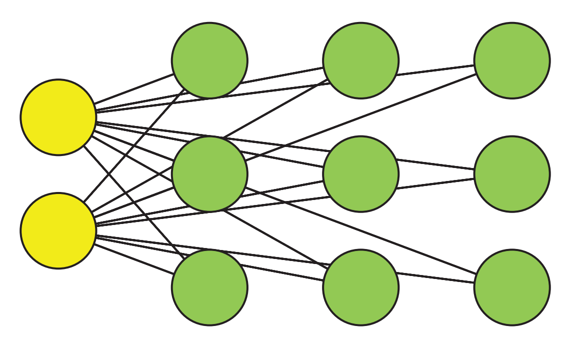
варианты нейросетей. Например, генеративно-состязательные нейронные сети нашли очень широкое применение в вопросе так называемого машинного творчества.

Финский учёный в области искусственного интеллекта Теуво Кохонен предложил несколько интереснейших концепций в рамках машинного обучения и искусственных нейронных сетей, из-за чего он стал наиболее часто цитируемым финским учёным. В частности, им предложены:

- фундаментальная теория ассоциативной памяти;
- особенный алгоритм обучения нейронных сетей;
- особый класс нейронных сетей (нейронные сети Кохонена);
- самоорганизующиеся карты Кохонена;
- модель нейрона и специального слоя (слой Кохонена).

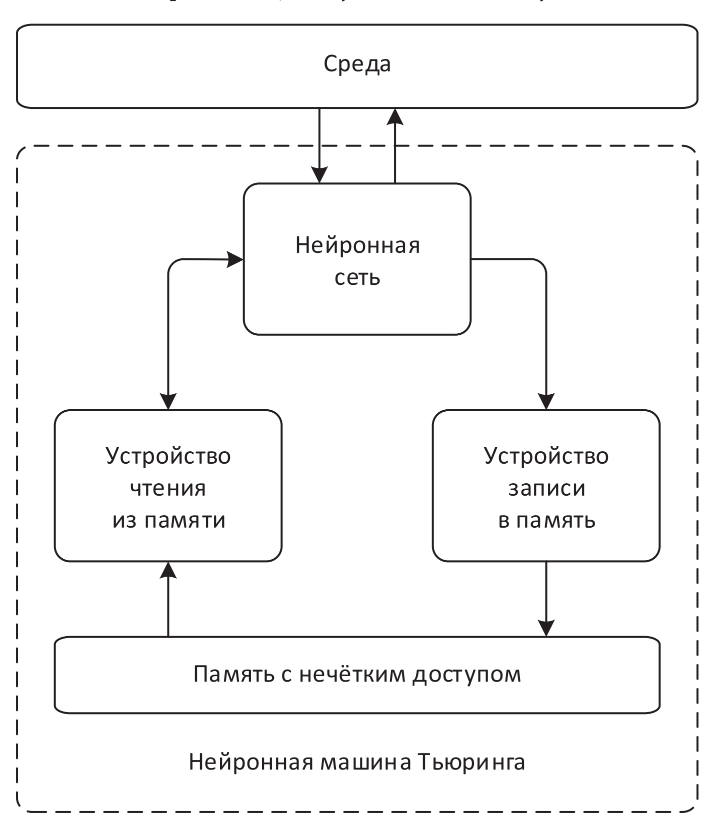
Наибольший интерес представляет самоорганизующаяся карта Кохонена. Это особая архитектура нейронной сети Кохонена для обучения без учителя. Карта решает задачи кластеризации и снижения размерности и применяется для решения задач моделирования, прогнозирования, выявления наборов независимых признаков, поиска закономерностей в больших массивах данных, квантизации признаков к их ограниченному числу индексов и некоторых других.

Самоорганизующаяся карта Кохонена получает на вход массив многомерных данных и проецирует его на двумерную плоскость, «раскрашивая» её и получая что-то вроде разноцветной административной карты территории (отсюда и название архитектуры). От разработчика или аналитика необходимо только задание количества кластеров, на которые необходимо разбить входной массив, а дальше нейросеть всё сделает самостоятельно.



Самоорганизующаяся карта Кохонена

Очень интересный подход предложили несколько учёных в 2014 г. В рамках него искусственную нейронную сеть совместили с внешней памятью, ведь произвольный доступ к памяти является критически важной функцией при обработке информацией. И компьютеры, и мозг человека работают с тем или иным видом памяти, к которой можно обращаться более или менее произвольно. Так вот, в предлагаемой архитектуре, которая была названа «нейронной машиной Тьюринга», нейронная сеть получает входную информацию и выдаёт выходную информацию не только во взаимодействии с внешним миром, но и использует внутреннюю память в виде матрицы чисел. Ячейки этой матрицы-памяти индексируются при помощи так называемых нечётких индексов, что позволяет нейросети обращаться ко всем ячейкам как бы одновременно, получая их линейную комбинацию.



На диаграмме показана общая структурная схема нейронной машины Тьюринга. Нейронная сеть является управляющим устройством машины и получает из среды входную информацию, выдавая в неё выходную информацию. Нейронная сеть управляет устройствами чтения и записи в память, которые, соответственно, осуществляют чтение и запись в память с нечётким доступом. Обращение к памяти осуществляется на каждом цикле работы нейронной сети, но при этом она сама обучается, как и когда это делать.

Эта архитектура, как предполагается, имеет много интересных приложений. Уже сегодня в экспериментах по распознаванию образов, обработке текста и по решению некоторых других задач такая нейронная машина Тьюринга показала результаты, превосходящие по производительности рекурсивные сети с LSTM-нейронами.

В большинстве моделей для кластеризации данных предполагается, что количество кластеров определяется заранее до старта алгоритма, сам алгоритм кластеризации работает один раз, а при изменении входных данных его необходимо запускать повторно (возможно, с изменением числа кластеров). Это схема работы большинства моделей машинного обучения, в том числе и многих нейросетевых моделей. В качестве альтернативной модели была предложена модель расширяющегося нейронного газа, в которой нет этих недостатков, – в ней число кластеров определяется самостоятельно, а сама модель является динамической, так что при изменении входных данных она тут же пересчитывает свои параметры и выдаёт новый результат, основанный на истории своей работы, в том числе и до изменения входных данных.

Существует ещё большое количество архитектур искусственных нейронных сетей, некоторые из которых уже устарели и представляют собой только архивно-исследовательский интерес, а некоторые только разработаны и ещё не до конца исследованы. Из наиболее интересных можно упомянуть следующие: когнитрон и неокогнитрон, осцилляторная нейронная сеть, вероятностная нейронная сеть Решетова, нейронная сеть адаптивного резонанса, импульсная нейронная сеть, нечёткая нейронная сеть и даже хаотическая нейронная сеть. Описания каждой из них можно найти в специальной литературе, а мы постепенно заканчиваем этот и без того раздувшийся раздел.

## Раздел 2.3. Эволюционные алгоритмы

Следующим классом методов, являющимся ещё одним представителем восходящей парадигмы создания искусственного интеллекта, являются эволюционные алгоритмы. Этот класс методов является отдельным направлением в рамках исследований по искусственному интеллекту, в котором исследуются и моделируются процессы естественного и искусственного отбора. Все эволюционные алгоритмы моделируют базовые эволюционные процессы в природе – наследование, мутации и отбор.

В применении к технологиям искусственного интеллекта работа эволюционных алгоритмов всегда связана с взаимодействием искусственной интеллектуальной системой с какой-либо средой, а опосредованно через неё, возможно, и с другими интеллектуальными системами (как искусственными, так и естественными). Такая среда сама может быть как искусственной (виртуальный мир, «песочница»), так и представлять собой объективную реальность, с которой искусственная интеллектуальная система взаимодействует при помощи набора сенсоров и исполнительных механизмов.

На текущий момент в рамках эволюционных алгоритмов выделяются следующие направления:

- эволюционное программирование;
- генетическое программирование;
- эволюционные стратегии;
- дифференциальная эволюция;
- генетические алгоритмы;
- нейроэволюция.

Необходимо отметить, что иногда некоторые исследователи располагают эволюционные алгоритмы (и даже более конкретно – генетические алгоритмы) в качестве одного из методов машинного обучения, входящего во множество вариантов обучения с подкреплением. Однако, по моему мнению, это отнесение генетических алгоритмов к машинному обучению в корне некорректно, и все эволюционные алгоритмы составляют отдельный класс методов восходящей парадигмы искусственного интеллекта.

Что, если вычислительные процессы могли бы эволюционировать так же, как это делают биологические виды в своей экологической среде? Возможно, получилось бы «выращивать» программы для оптимального решения поставленной задачи? Эволюционное программирование как раз и обратилось к этой идее, когда

в 1960 г. доктор Лоуренс Фогель предложил рассматривать «разумные системы» как такие, которые способны моделировать окружающую их среду и делать на основании этого те или иные прогнозы для наиболее оптимального достижения своей цели. Фактически это было началом агентного подхода в искусственном интеллекте.

Такие «разумные системы» предлагалось оценивать и отбирать на основании наилучших оценок. Фактически речь шла о программах, которые получали на вход некоторые данные для обработки и возвращали результат. И вот подбор осуществлялся именно на входных данных. В реальности получился один из эвристических методов оптимизации, в котором движение к целевому состоянию системы осуществлялось при помощи максимизации выходной функции на основе варьирования входных параметров. И в этом случае оптимизация как бы моделирует эволюционный процесс.

А что, если применить эволюционные принципы на текстах программ? Пусть программы проходят отбор на то, что результаты их выполнения соответствуют заданным критериям останова. Если вспомнить, что программы – это те же символьные записи, которые могут быть закодированы при помощи битов, то особых различий с предыдущим вариантов эволюционных алгоритмов нет. Достаточно лишь программу воспринимать в качестве входных данных. Тем более что сам этот принцип уже многократно использовался –и универсальная машина Тьюринга принимает описание программ других машин Тьюринга в качестве входных данных, и язык программирования ЛИСП воспринимает в виде входных данных исходные тексты своих программ.

Дополнительным условием в этом случае будет являться корректность с точки зрения синтаксиса конкретного языка программирования. В рамках корректно построенных программ их можно подвергать эволюционному процессу для нахождения наиболее оптимальных программ, выполняющих заданную функцию. Именно в этом заключается суть генетического программирования.

Далее. Эволюционная стратегия – это эвристический метод оптимизации, разработанный в 1964 г. Инго Рехенбергом. Фактически этот метод стал предтечей генетических алгоритмов, хотя в том или ином виде они прорабатывались и ранее. В эволюционной стратегии осуществляется поиск целевого вектора, компонентами которого являются действительные числа. Для поиска решения осуществляется скрещивание особей (векторов) и мутации. Далее применяется функция отбора, которая принимает на вход всю ко-

горту: родительские особи и их потомков. Отбор осуществляется детерминированным образом – оставляются самые лучшие особи, причём без повторений. Операция мутации может быть произвольной над действительными числами, но обычно используется простое добавление нормально распределённого случайного числа к компонентам векторов. Важным дополнением эволюционной стратегии является то, что в процессе эволюции параметры нормального распределения добавляемой в рамках мутации случайной величины адаптируются под поиск решения задачи.

Дифференциальная эволюция – тоже метод оптимизации, который позволяет найти глобальный экстремум для недифференцируемых, нелинейных, мультимодальных функций от многих переменных. Это очень вычислительно простой и в то же время достаточно мощный метод, позволяющий довольно быстро найти именно глобальный экстремум.

Суть его проста. Пространство поиска представляет собой векторы. Для каждого вектора текущего поколения случайным образом выбираются три других вектора из этого же поколения, над которыми производится операция мутации, являющаяся параметром алгоритма. Например, к первому вектору добавляется удвоенная разность между вторым и третьим векторами. Далее полученный мутантный вектор подвергается «слиянию» с изначальным вектором – для каждого элемента изначального вектора производится случайный выбор того, замещать ли его соответствующим элементом мутантного вектора. После этой операции получается пробный вектор, на котором вычисляется значение фитнесс-функции. Если оно ближе к экстремуму, чем значение на изначальном векторе, то в новое поколение идёт пробный, а иначе в нём остаётся изначальный. И такой процесс осуществляется над каждым вектором текущего поколения.

Перейдём к наиболее известному из методов, который является ярким представителем эволюционного подхода. Это генетические алгоритмы. Генетические алгоритмы сами по себе опять являются одним из эвристических методов оптимизации для поиска оптимального решения (или, как минимум, субоптимального). Они работают с данными, которые могут быть представлены в виде «хромосом» – последовательностей генов, т. е. списков каких-либо значений, к которым можно применить генетические операции. Здесь уже нет такого жёсткого требования насчёт типа таких данных, главное – чтобы на генах были определены эти операции, которые возвращали бы приемлемый результат, имеющий смысл.

Генетический алгоритм состоит из следующих шагов.

1. Генерация начальной популяции.
2. Цикличный процесс рождения новых поколений и отбора.
3. Остановка и возвращение результатов поиска.

Итак, генерация начальной популяции. На этом шаге необходимо подготовить некоторое количество «начальных» значений в пространстве поиска, с которых алгоритм начнёт свою работу. Начать можно с произвольных значений, но если они будут как можно более близки к целевому, то алгоритм отработает намного быстрее.

Далее – запуск цикличного процесса рождения новых поколений и отбора. Фактически это и есть сам генетический алгоритм, который раз за разом запускает процесс порождения новых поколений, изучения новых особей и отбора наиболее интересных. В свою очередь, этот процесс состоит из следующих шагов и операций: скрещивание, мутация и отбор.

Операция скрещивания, или перекрёста (*англ.* crossingover), – это операция, при которой берутся две особи текущего поколения и их генотип перемешивается друг с другом. Чаще всего два генотипа разрезаются в одинаковом случайном месте, и после этого вторые части переставляются друг с другом. Так получается два новых потомка. Фактически скрещивание делается каждой особи с каждой, да ещё и несколько раз, чтобы разрезание генотипа происходило в разных местах. Так получается набор особей следующего поколения, при этом обеспечивается генетическое разнообразие.

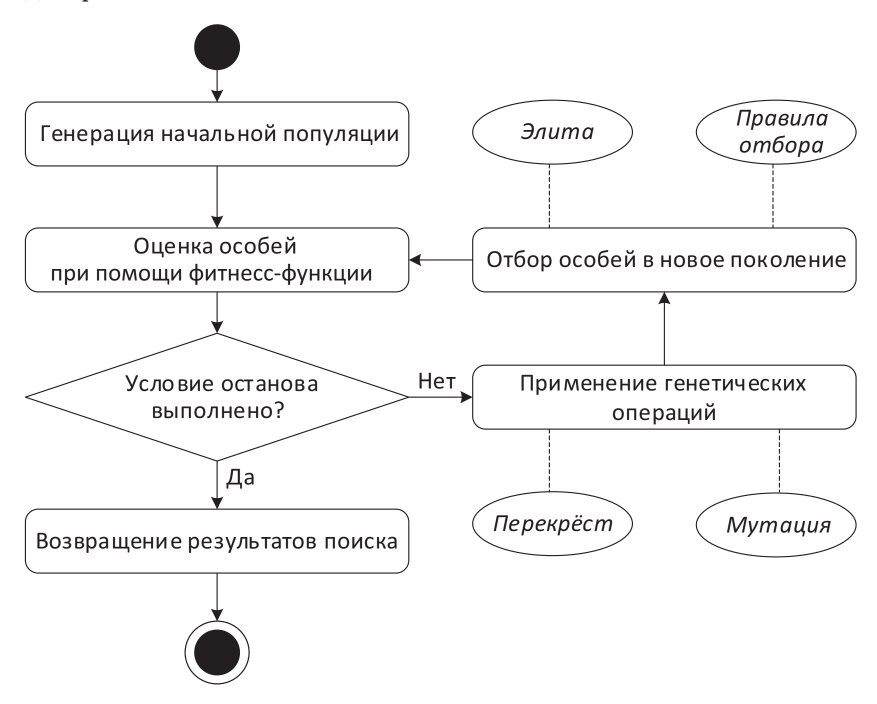
Операция мутации – это внесение случайных изменений в генотип потомства. С установленной частотой в каждый новый генотип после скрещивания вносятся случайные изменения. Если частота мутаций высокая, то наблюдается «подвижность» поиска, если же частота мутаций слишком низкая, то часто это приводит к «залипанию» алгоритма в локальном экстремуме, из которого он не может выбраться и, соответственно, остановиться. Величину частоты мутаций подбирают при помощи экспериментов для каждой конкретной задачи.

Операция отбора – это выбор наиболее приспособленных особей. Обычно выбирается какая-то небольшая доля (например, 10%) тех особей нового поколения, для которых фитнесс-функция возвращает наилучшие результаты. Иногда в отобранное множество особей нового поколения добавляются наименее приспособленные особи, для которых фитнесс-функция возвращает наихудшие результаты. Это делается для получения генетического разнообразия и является одним из инструментов борьбы

с «залипанием» в локальных экстремумах. Также иногда в состав нового поколения включаются некоторые представители предыдущего поколения, которые в этом случае называются «элитой». Вообще, операция отбора подбирается на основе экспериментов для каждой конкретной задачи.

И наконец, третий шаг – возвращение результатов, которые привели к остановке цикла. Завершение алгоритма происходит тогда, когда фитнесс-функция для какой-либо особи находится в окрестности целевого значения с заданной точностью. Собственно предикат проверки условия для остановки алгоритма проверяет для каждой особи значение фитнесс-функции, и если абсолютное значение разницы между ним и целевым значением меньше заданной точности, то предикат возвращает истину, и алгоритм останавливается. Необходимая точность является входным параметром алгоритма.

Общая схема генетического алгоритма показана на следующей диаграмме.




Собственно, любую задачу, которую можно решить при помощи генетического алгоритма, можно уложить в эту схему. Это

значит, что сам алгоритм может быть реализован в виде обобщённой функции высшего порядка, которой на вход передаётся набор функциональных и обычных параметров:

- функция генерации случайной особи;
- фитнесс-функция;
- функция перекрёста, включая количество разрезаний хромосомы;
- функция мутации, включая частоту мутации;
- функция отбора, включая все условия попадания особи в новое поколение – есть ли элита, какой процент особей сверху и снизу отбирается в новое поколение;
- количество особей в поколении;
- требуемый уровень точности;
- целевое значение фитнесс-функции.

Забавно то, что в этот алгоритм можно «уложить» и подбор параметров для методов машинного обучения и даже для самих эволюционных алгоритмов. Дело в том, что методы восходящего подхода искусственного интеллекта несколько несовершенны, а потому часто используются в автоматизированном режиме, когда исследователь подбирает параметры обучения или генетического алгоритма вручную, изучая результаты прогонов алгоритмов на тех или иных значениях параметров. Вот так и здесь – параметры генетического алгоритма можно подвергнуть подбору при помощи генетического алгоритма.

Если взять числовые параметры генетического алгоритма, перечисленные ранее, – количество разрезаний хромосомы, частоту мутаций и т. д., – то из них можно составить вектор, т. е. хромосому, которая и подвергается генетическому алгоритму. Фитнес-функцией в этом случае является функция, оценивающая скорость схождения алгоритма и отсутствие залипания в локальных экстремумах.

И наконец, очень интересная и важная идея – применение эволюционных алгоритмов к нейронным сетям. А что, если в виде хромосомы представить нейронную сеть с настройками и параметрами обучения, а потом над некоторой начальной популяцией таких нейронных сетей осуществить процедуру дифференциальной эволюции или генетического алгоритма? В этом и состоит суть *нейроэволюции*. И сегодня это одно из самых интересных направлений в рамках восходящего подхода. Искусственный интеллект начинает использовать эволюцию для создания ещё более мощного искусственного интеллекта. Так победим.

# Глава 3

# СОВРЕМЕННОЕ СОСТОЯНИЕ ТЕХНОЛОГИЙ

Попробуем детально изучить современное состояние технологий искусственного интеллекта. Следует отметить, что на сегодняшний момент можно констатировать окончание второй из так называемых «зим искусственного интеллекта», поскольку большинство направлений получило новые усилия в своём развитии. Так что очень многие технологии именно в настоящий момент развиваются семимильными шагами, поэтому приведённая в этой главе информация может оказаться устаревшей очень быстро. Тем не менее сделаем некоторый срез на 2017–2019 гг. и рассмотрим следующие направления, которые уже были представлены в виде типовых задач в первой главе:

- машинное обучение;
- распознавание образов;
- интеллектуальный анализ данных, или «дата-майнинг»;
- обработка НЕ-факторов знания;
- принятие решений;
- поиск информации;
- обработка естественного языка;
- представление знаний;
- робототехника;
- роевой интеллект.

Далее кратко рассматривается каждое из перечисленных направлений, даётся его характеристика, доступно и достаточно для понимания приводится формальная сторона вопроса, описываются современные достижения, а также в некоторых случаях перечисляются направления будущих исследований и открытые вопросы.

Необходимо отметить, что целью этой книги не является детальное погружение в технологии искусственного интеллекта, поэтому в данной главе приводится лишь описание сути технологий. За деталями читатели направляются в специализированную литературу и на MOOC-курсы.

## Раздел 3.1. Методы поиска

В рамках искусственного интеллекта выделяется задача поиска, которая представляется крайне важной. Фактически поиск в довольно широком смысле – это нахождение траектории в некотором пространстве от текущего состояния к целевому. Это очень общее определение, так как и пространство поиска, и целевая функция могут быть произвольными. Для искусственного интеллекта очень важно уметь осуществлять поиск, поскольку поиск – это один из методов познания реальности, получения новых знаний. Так что в этом разделе имеет смысл рассмотреть некоторые особо важные технологии и задачи поиска в искусственном интеллекте.

Выделяются следующие типы поиска, которые имеют непосредственное отношение к методам искусственного интеллекта:

- поиск в пространстве состояний;
- поиск пути;
- эвристический поиск;
- информационный поиск.

Эти типы поиска не являются взаимоисключающими. Это именно типы поисковых алгоритмов, и некоторые из таких алгоритмов, к примеру, могут осуществлять информационный поиск при помощи эвристик случайного блуждания в пространстве состояний. Так что имеет смысл рассмотреть каждый тип, но принимая во внимание то, что некоторые алгоритмы могут обладать свойствами нескольких типов.

Поиск в пространстве состояний осуществляется для нахождения траектории системы в фазовом пространстве её состояний из некоторого начального состояния до целевого. Целевое состояние определяется при помощи предиката, реализующего алгоритм проверки, является заданное состояние целевым или нет (очевидно, что целевых состояний может быть несколько, и предикат описывает ту желательную конфигурацию системы, в которую она должна попасть в результате поиска). Поиск осуществляется при помощи использования двух функций – определения следующих возможных состояний для заданного и определения стоимости перехода из одного состояния в другое.

Фактически такая постановка задачи позволяет описать задачу поиска при помощи теории графов, когда состояниям системы

соответствуют вершины графа, направленные рёбра графа, – это возможность перехода из одного состояния в другое, а функция стоимости пути определяет веса на рёбрах. В таком представлении предикат целевых состояний определяет те вершины графа, в которые необходимо перейти из текущей. А функция стоимости пути, соответственно, позволяет минимизировать (если это возможно) сумму всех весов на обходимых рёбрах.

Из такой постановки задачи поиска в пространстве состояний понятно, что это задача оптимизации. А из этого с очевидностью следует, что решать её можно различными методами, имеющимися в арсенале теории оптимизации для решения задач на графах.

Для решения этой задачи разработано достаточное количество методов и алгоритмов, которые разделяются на информированные и неинформированные. Информированные методы поиска пользуются различными эвристиками, определяющими дополнительную информацию о той задаче, которая решается. Это часто позволяет сократить перебор состояний системы и оптимизировать поиск. Но многие информированные методы зачастую применимы только для конкретных задач. В отличие от них, неинформированные методы поиска, также называемые методами грубой силы, пользуются только информацией о том, как отличить целевое состояние от нецелевых. Фактически различие в методах этой группы сводится к порядку перебора состояний в процессе поиска.

Неинформированные методы поиска подразделяются на следующие варианты.

- *Поиск с возвратом* – фактически универсальный метод поиска, осуществляющий полный перебор вариантов. Алгоритм позволяет найти *все* решения задачи, если они существуют, однако время его работы недетерминированно, и в случае гигантских размеров пространства поиска практическое применение метода может быть невозможным.
- *Поиск в ширину* (BFS) – поиск, в котором из начальной вершины графа состояний просматриваются все её смежные вершины, а переход в следующую вершину происходит только после просмотра всех вершин текущего уровня. Если любой просмотр имеет одинаковую стоимость, то данная стратегия поиска является оптимальной. В реальных задачах этого добиться непросто.

- *Поиск по критерию стоимости* (UCS) обобщает предыдущий метод на случай, когда переход в разные смежные вершины в графе состояний имеет разную стоимость. На каждом шаге просматривается смежный узел, переход в который имеет наименьшую стоимость. Эта стратегия оптимальная в том случае, если все стоимости строго положительны.
- *Поиск в глубину* (DFS) – такой поиск, при котором переход в пространстве состояний осуществляется до самой глубокой вершины в графе, а потом поиск возвращается на шаг назад или ещё ранее, в случае если целевой вершины не обнаружено, а смежных непросмотренных вершин у текущего узла графа больше нет.
- *Поиск с ограничением глубины* (DLS) – вариант поиска в глубину, который ограничивает глубину просмотра некоторым заданным параметром, что позволяет решить проблему бесконечного пути, на котором «спотыкается» предыдущая стратегия.
- *Поиск в глубину с итеративным углублением* (IDDFS) – расширение предыдущей стратегии, при котором заданное максимальное значение глубины постепенно увеличивается до тех пор, пока не будет найдено целевое состояние.
- *Двунаправленный поиск* – более сложная стратегия как для поиска в ширину, так и для поиска в глубину, когда известна целевая вершина и необходимо найти путь из начальной в целевую. В этом случае можно применять ту или иную стратегию, запустив процессы поиска навстречу друг другу из начальной в целевую и из целевой в начальную.

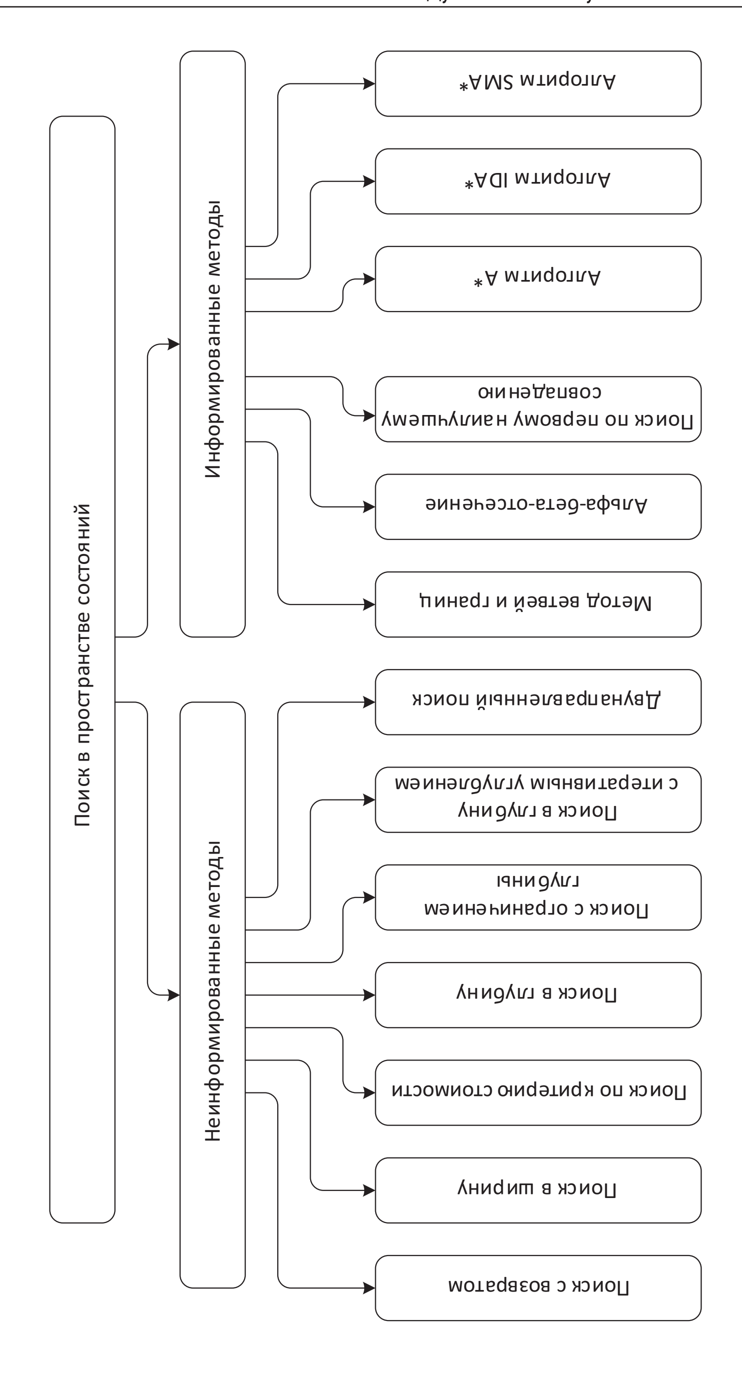
В свою очередь, информированный поиск может вестись при помощи следующих алгоритмов.

- *Метод ветвей и границ* – это общий метод для оптимизации, являющийся вариантом метода полного перебора с отсевом подмножеств допустимых решений, которые заведомо не содержат оптимального.
- *Альфа-бета-отсечение* – алгоритм поиска, стремящийся сократить количество узлов, оцениваемых в дереве поиска алгоритмом минимаксной оптимизации. В основе алгоритма лежит идея, что оценивание ветви дерева поиска может быть досрочно прекращено без вычисления всех значений оценивающей функции, если было найдено, что для этой ветви значение оценивающей функции в любом случае хуже, чем вычисленное для предыдущей ветви.

- *Поиск по первому наилучшему совпадению*. Если имеется правило оценки заданной вершины, то алгоритм поиска по первому наилучшему совпадению исследует граф состояний системы, выбирая для расширения вершины в соответствии с этим правилом оценки. Фактически это общий вид эвристического поиска, когда правило оценки задаётся функцией эвристики, принимающей во внимание не только саму вершину, но и любое количество иных факторов (например, вся историческая информация, собранная алгоритмом поиска к текущему моменту).
- *Алгоритм A** – вариант предыдущего метода, наиболее часто используемый для информированного поиска по графу. В качестве эвристической функции используется функция «расстояние + стоимость», где расстояние от текущей вершины до целевой оценивается эвристически, а стоимость определяется функцией достижения текущей вершины из начальной, и эта функция может быть как эвристической, так и нет.
- *Алгоритм IDA** – определённого рода сочетание алгоритма A* и алгоритма итеративного углубления (метод неинформированного поиска). Алгоритм может остановить развёртывание вершины в двух случаях: когда глубина поиска превышает установленный предел и когда оценка стоимости пути через текущую вершину превышает текущий предел стоимости, т. е. используются два критерия из обоих алгоритмов.
- *Алгоритм SMA** – упрощение алгоритма A*, работающее в ограниченной памяти, в то время как алгоритм A* может требовать экспоненциального объёма памяти для своей работы. Единственное отличие – когда вся отведённая для работы алгоритма память использована, то из списка вершин для посещения удаляются вершины, имеющие наибольшее значение стоимости. Все остальные шаги такие же, как в алгоритме A*.

Конечно, здесь перечислены не все методы и алгоритмы поиска, разработанные к настоящему времени, но основные классы таких методов. В дополнительных источниках можно найти описания более специализированных методов поиска обоих классов.

Таким образом, основные методы поиска в пространстве состояний можно классифицировать следующим образом.


Поиск в пространстве состояний
Информированные методы
Неинформированные методы
Алгоритм SMA*
Алгоритм IDA*
Алгоритм A*
Поиск по первому наилучшему совпадению
Альфа-бета-отсечение
Метод ветвей и границ
Двунаправленный поиск
Поиск в глубину с итеративным углублением
Поиск с ограничением глубины
Поиск в глубину
Поиск по критерию стоимости
Поиск в ширину
Поиск с возвратом

Что ж, перейдём к следующему типу поиска – поиску пути. Этот тип поиска крайне востребован в таких областях, как игровой искусственный интеллект и навигация, причём вторую область можно разделить на навигацию для человека (т. е. построение оптимального маршрута от начальной точки до конечной с учётом возможных промежуточных) и для автономного интеллектуального агента, типа автономного автомобиля (фактически то же самое, но с учётом дополнительных ограничений и нюансов функционирования такого агента). По своей сути алгоритмы поиска пути являются теми же самыми алгоритмами поиска на графах, только теперь граф представляет не отображение перехода между состояниями системы, а более конкретизированную вещь – отображение локаций и путей между ними в физической или виртуальной реальности.

Если граф пространства, в котором ищется путь, не слишком сложный, то можно воспользоваться алгоритмами поиска в пространстве состояний. Обычно для этих целей используется алгоритм A* с различными эвристиками, подбираемыми под конкретную задачу. Если пространство (географическая местность или виртуальное пространство) более сложное, то резонно воспользоваться специализированными алгоритмами. В качестве интересных примеров таких алгоритмов можно отметить следующие.

- *Алгоритм Дейкстры* и его расширения. Алгоритм находит кратчайшие пути от заданной вершины до всех остальных. Работает только для графов, в которых нет рёбер с отрицательной стоимостью, однако такие его расширения, как алгоритм Беллмана–Форда, алгоритм Джонсона и алгоритм Левита, работают и для графов с наличием отрицательных рёбер. Но у каждого есть свои ограничения. В процессе своей работы алгоритм обходит все вершины графа и каждой из них приписывает метку – текущее минимальное расстояние до этой вершины от заданной. Когда алгоритм обойдёт все вершины, работа заканчивается, а каждой вершине сопоставлена метка, равная минимальному расстоянию.
- *Волновой алгоритм*. Алгоритм, который предназначен для поиска пути на *дискретном рабочем поле*, представляющем собой ограниченное пространство, разбитое на прямоугольные ячейки. Осуществляет поиск кратчайшего пути из стартовой ячейки до целевой, если такой путь существует. Разбивается на три этапа – инициализацию ячеек, распространение волны и восстановление пути. Имеет два

варианта с точки зрения связности ячеек – ортогональный и ортогонально-диагональный. Может быть обобщён на ячейки произвольной формы, даже не совпадающие по форме в рамках одного дискретного рабочего поля. Чаще всего применяется в компьютерных играх и виртуальной реальности для поиска пути в пространстве. Кроме того, эффективно используется для трассировки печатных плат.

- *Муравьиный алгоритм*. Довольно эффективный алгоритм оптимизации для нахождения приближённых решений задачи по обнаружению самого выгодного маршрута, проходящего через указанные узлы графа хотя бы по одному разу с последующим возвратом в исходную вершину (задача коммивояжёра). В основе алгоритма лежит поведение муравьёв-фуражиров, которые собирают для колонии корм. Фактически это один из первых алгоритмов восходящей парадигмы искусственного интеллекта, поскольку моделирует низкоуровневую биологическую систему на базе многоагентного подхода. Муравьи изначально ходят случайно вокруг муравейника, но после обнаружения источника пищи они прокладывают тропу при помощи феромонов. Феромоны имеют тенденцию испаряться, если их не обновлять, и это значит, что более короткие маршруты имеют преимущество перед длинными. Испарение феромонов также играет важную роль для предотвращения попадания в локальные экстремумы.
- *Методы динамического программирования*. Основная суть динамического программирования заключается в декомпозиции задачи на подзадачи и применении к упрощённым задачам тех же методов решения. Если декомпозиция возможна, то задача разбивается на настолько дробные подзадачи, насколько это осуществимо, и тогда каждая мелкая подзадача, возможно, решается более просто. После решения каждой подзадачи собирается окончательный вариант. Этот метод не является универсальным, и в каждом случае необходимо смотреть, есть ли возможность его применения. Кроме того, это как бы метаметод, метод высшего порядка, так как в его рамках могут использоваться любые другие методы для решения атомарных подзадач.

Третий тип поиска – эвристический. Вообще, эвристика – это некоторый не совсем точный и практический алгоритм решения некоторой задачи, который не гарантирует точность или

оптимальность решения, но вполне достаточен для достижения практического результата. Эвристики ускоряют решение задач там и тогда, когда точное аналитическое решение не может быть найдено или с практической точки зрения, или из-за фундаментальных ограничений. Другими словами, эвристический алгоритм является не полностью обоснованным математически, а иногда и вовсе некорректным, но при этом вполне полезным.

Эвристики обычно применяются там, где используются задачи высокой вычислительной сложности, при этом оценка стоимости применения точных алгоритмов (которых может и не существовать в принципе) существенно выше, чем стоимость применения эвристического алгоритма. Само собой разумеется, что эвристика и возможность её применения подбирается для каждой конкретной задачи. Выше уже было показано, как эвристики использовались в других методах поиска. Но при этом эвристические алгоритмы вполне применимы и в других областях, в том числе тех, которые рассматриваются далее, например в задачах распознавания образов, машинного вывода, машинного творчества.

Наконец, перейдём к информационному поиску. Нельзя сказать, что это чистая технология искусственного интеллекта. Скорее, это вспомогательная технология, которая используется искусственными интеллектуальными системами для получения дополнительной информации. Особенно если речь идёт о символьных или гибридных системах. Другими словами, информационный поиск, развившийся из первых поисковых машин в интернете, сегодня становится неотъемлемой частью серьёзных интеллектуальных систем, поскольку такие системы, для того чтобы показывать высокую эффективность, должны иметь доступ ко всему корпусу знаний, накопленному к настоящему моменту человечеством.

Обычно задача информационного поиска ставится как выявление информации, удовлетворяющей с той или иной степенью поисковому запросу, из неструктурированных документов. Чаще всего поисковый запрос формализован, а вот источники информации представлены в виде неформализованных и несвязанных документов. Соответственно, с точки зрения искусственной интеллектуальной системы речь идёт об извлечении знаний из источников третьего рода.

Информационный поиск можно разделить на следующие виды.

1. *Поиск по метаданным* – это наиболее простой вид информационного поиска, здесь достаточно использования самых

простых алгоритмов, реализованных в системах управления базами данных. Поиск ведётся по полностью формализованной и структурированной информации на основе запроса на каком-либо формальном языке запросов.

2. *Полнотекстовый поиск* – наиболее сложный вид информационного поиска, задача которого в общем виде не решена до сих пор. Поиск ведётся по неструктурированным текстам, и чаще всего требуется найти нечто, отвечающее определённому смыслу. Это задача более высокого порядка, нежели простой поиск соответствия в корпусе текстов какой-либо входной строки (информационно-поисковые машины). Система, осуществляющая полнотекстовый поиск, должна распознать смысл входного запроса и дать ответ в соответствии с этим смыслом. Многие методы искусственного интеллекта – как символьные, так и «грязные» – направлены именно на решение этой задачи.
3. *Поиск по изображениям*. Обычно эта задача решается при помощи методов машинного обучения, и в последнее время в связи с возросшими вычислительными мощностями и появлением архитектур глубокого обучения она решается вполне успешно. Искусственная интеллектуальная система может найти на изображении заданную информацию, выделить лица или животных, отделить фон от изображений первого ряда и т. д.

Итак, информационный поиск является неотъемлемой частью искусственных интеллектуальных систем, в том числе и потому, что в процессе своей работы они должны постоянно получать актуальную информацию в области своей деятельности. Это вопрос уже постоянного обучения системы, будь она построена в рамках нисходящей или восходящей парадигмs. А для осуществления информационного поиска чаще всего требуются технологии обработки естественного языка, так как большая часть знаний человечества записана именно в неформальном, неструктурированном виде.

## Раздел 3.2. Обработка естественного языка

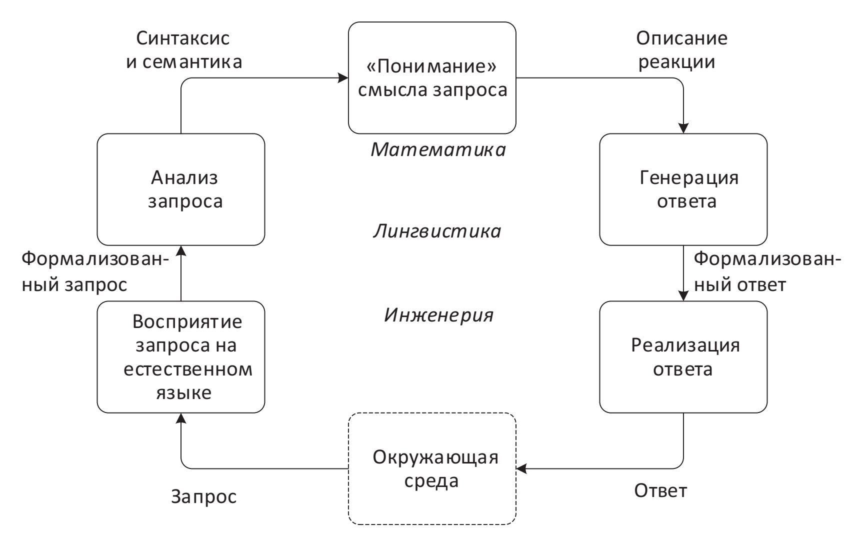
Обработка естественного языка – это довольно большой раздел искусственного интеллекта, лежащий на пересечении интересов математики, лингвистики и прикладных инженерных наук.

Основная задача заключается в возможности интеллектуальной системы осуществлять анализ и синтез фраз на естественном языке для общения с человеком. Вопрос о том, «понимает» ли эта интеллектуальная система то, о чём ведётся диалог, в рамках этого направления не ставится, так как человек сам ещё не очень-то понимает, что значит «понимать». Это вопрос философии сознания, а обработка естественного языка решает вполне прикладную задачу.

Процесс обработки естественного языка можно представить в виде кибернетической диаграммы, состоящей из следующих шагов:

- восприятие запроса на естественном языке;
- анализ запроса;
- «понимание» смысла запроса;
- генерация ответа;
- реализация ответа.

Сама диаграмма выглядит следующим образом.




На ней очень хорошо видны уровни применения пересекающихся в рамках технологий обработки естественного языка наук.

Итак, интеллектуальная система взаимодействует со своим окружением при помощи естественного языка. Как же происходит коммуникативный акт? Его можно описать следующим образом.

1. Окружающая среда формирует для интеллектуальной системы сообщение на естественном языке. Это может быть фраза, высказанная голосом и принятая сенсорами системы в виде акустических волн. Это может быть фраза, записанная на бумаге письменным курсивом и переданная в систему уже в виде матрицы пикселей. А также это может быть напечатанная на клавиатуре фраза, и тогда она непосредственно передаётся в систему.
2. Любой вариант входной естественно-языковой фразы попадает с сенсора на первичный обработчик, который и осуществляет процесс восприятия запроса на естественном языке. Фактически задачей этого этапа является перевод фразы из входного формата в формальное представление в виде последовательности символов – букв, слов, предложений. И это чисто инженерная задача (хотя, конечно же, она также нередко решается методами искусственного интеллекта, в частности нейросетевые модели в этой задачи получили самое широкое распространение и применение).
3. Далее формализованное сообщение попадает в лингвистический процессор, где осуществляется анализ сообщения для выявления его смысла в контексте коммуникативного акта. Лингвистический процессор сам по себе является довольно сложной системой, состоящей из последовательно соединённых блоков и использующей многостороннюю лингвистическую информацию и информацию об объективной реальности. Лингвистический процессор позволяет на основе входной фразы получить варианты дальнейших действий интеллектуальной системы, в состав которой он входит. Задачами лингвистического процессора являются: выявление и исправление ошибок во входной фразе, её токенизация, морфологический анализ каждого токена, синтаксический и семантический анализы всего сообщения. Предназначением же лингвистического процессора в рамках использующей его интеллектуальной системы является преобразование формализованной входной фразы в некоторый внутренний язык представления смысла, который система может перевести в последовательность своих поведенческих актов.
4. Проблема заключается в том, что естественный язык и общение на нём используют очень много неопределённостей,

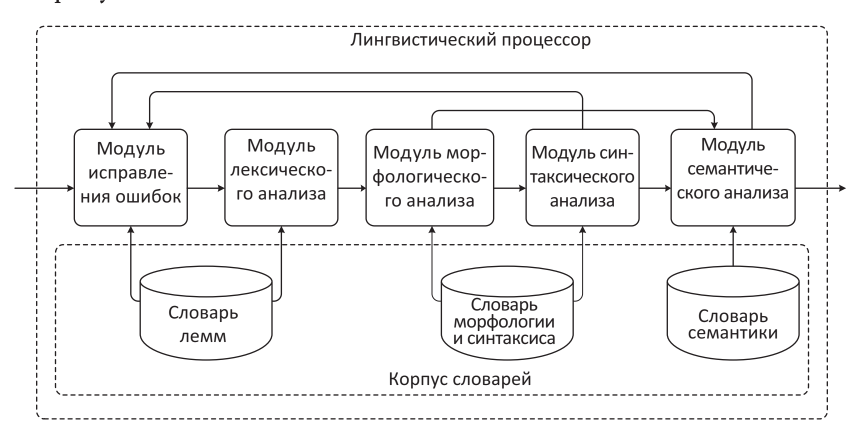
неоднозначностей, умолчаний и других подобных факторов. Даже человеческий естественный интеллект не всегда может правильно распознать смысл обращённой к нему фразы. Человеку в этом отношении помогают контекст беседы и общие знания об окружающей реальности. То же самое должна иметь и интеллектуальная система – в её базе знаний должно быть достаточно информации о том, как устроен окружающий мир (как минимум, в части той проблемной области, в которой ведётся разговор), и при этом система должна владеть контекстом беседы. Более того, интеллектуальная система также должна производить и так называемый прагматический анализ, т. е. изучать взаимоотношение и взаимодействие входной коммуникационной информации по отношению к самой себе и своему состоянию. Все это позволяет выбрать из полученных на выходе лингвистического процессора вариантов смысла один – именно тот, который подразумевался собеседником. На основании выбранного варианта смысла входной фразы интеллектуальная система осуществляет поведенческий акт, который на выходе из подсистемы «понимания» смысла запроса имеет вид формального описания того, что необходимо выполнить.

5. Следующий шаг заключается в генерации ответа на естественном языке. Этот шаг не является необходимым для интеллектуальных систем, от которых не ожидается ответа на запросы, но такие классы систем, как вопрос-ответные системы, чат-боты или иные системы с использованием общения с пользователем на естественном языке, должны осуществлять этот этап работы. В любом случае в состав методов обработки естественного языка, входят алгоритмы синтеза текста, но они намного проще, чем алгоритмы анализа. Перевод поведенческого акта системы с внутреннего языка представления смысла на естественный может осуществляться по шаблонам или по достаточно простым правилам генерации.
6. Наконец, последним шагом опять осуществляется в целом инженерная задача по переводу полученного естественно-языкового ответа в ту форму представления, которая доступна собеседнику. Например, если общение осуществляется посредством голоса, то интеллектуальная система должна сгенерировать звуковое соответствие своего ответа, т. е. как бы прочитать то, что написано. Ответ может пере-

даваться в виде радиоволн или ещё как-то. Чаще всего сгенерированный ответ в виде строки непосредственно передаётся в канал, при помощи которого происходит общение.

Наиболее важными и сложными являются шаги 2 и 3, поэтому заострим на них внимание.

Второй шаг выполняется лингвистическим процессором, который является программным средством, состоящим из нескольких модулей и многочисленных языковых баз данных. Его типовая структура может быть схематично представлена так, как показано на рисунке.




Кратко рассмотрим каждый компонент.

1. Модуль исправления ошибок осуществляет первичный просмотр входной естественно-языковой фразы и отлавливает в ней слова, которые могут быть написаны с ошибками. Это довольно простая задача в том смысле, что она легко формализуется и на её выходе неоднозначность минимальна. При этом не секрет, что если подходить к решению этой задачи «в лоб», то есть, например, для каждого входного слова возвращать список возможных замен для него (изменение одной буквы, перестановки букв и прочие варианты ошибочного написания), то общая неопределённость будет крайне высокой, так как практически у каждого слова или даже формы слова есть возможность изменить одну или несколько букв и получить другое правильное слово языка. Однако современные алгоритмы вполне понимают, когда имеется ошибка, а когда нет. К тому же часто алгоритмы этого модуля

реализуют две тонкости. Во-первых, вместе с входной фразой для проверки они могут принимать дополнительную метаинформацию о фразе, в том числе и о том, как она была получена. Например, при получении фразы с клавиатуры ошибки, связанные с опечатками (нажата соседняя кнопка на клавиатуре), имеют больший вес, нежели те, которые с опечатками не связаны (например, замена букв осуществлена кнопками, находящимися на разных сторонах клавиатуры). Во фразе, пришедшей после распознавания рукописного или печатного текста, напротив, больше внимания будет уделено тем ошибкам, которые связаны с заменами похожих по начертанию букв. Второй аспект таких алгоритмов связан с получением обратной связи от модулей морфологического, синтаксического и семантического анализов, на основании результатов работы которых также можно определить, имеются во входной фразе ошибки или нет.

2. Входная строка с пометками о потенциальны ошибках далее подаётся на вход в модуль лексического анализа (или «токенизации»), т. е. разделения на отдельные слова или словосочетания (иногда единицей смысла выступает не слово, а их сочетание, например «в связи с», «так же, как и» и др.). Фактически для каждого варианта входной фразы с исправленными ошибками на выходе модуля токенизации появляется список лексем, который далее подаётся на вход модулю морфологического анализа.
3. Модуль морфологического анализа принимает на вход список вариантов лексем входной фразы и при помощи морфологического словаря для каждого варианта лексемы возвращает список возможных вариантов морфологического разбора этой лексемы. Под этим понимается возвращение начальной формы лексемы и приписывание к ней всей морфологической информации – тех грамматических категорий, в которых находится изначальная форма лексемы. Для различных языков набор грамматических категорий различен, но для целей универсального морфологического анализа чаще всего используются такие категории, как синтаксический падеж, число, род – для имён существительных и имён прилагательных; вид, наклонение, лицо, время – для глаголов и т. д. Набор значений каждой грамматической категории также различается от языка к языку. В ито-

ге далее на вход модуля синтаксического анализа подаётся список вариантов лексем, и при этом к каждому варианту приписан список возможных способов морфологического анализа соответствующего варианта. Уже это позволяет понять, как в процессе лингвистического анализа расширяется пространство возможностей.

4. Следующим шагом в разборе фразы на естественном языке осуществляется синтаксический анализ, который выполняется соответствующим модулем. На его вход подаётся результат работы модуля морфологического анализа, и функциональность синтаксического анализа пытается сузить пространство возможностей при помощи применения ко всем вариантам морфологии лексем входной фразы различных синтаксических схем. В итоге оказывается, что не каждый вариант морфологии может использоваться в том окружении, в котором находится. По результатам работы отсекается огромная часть возможностей, и на выходе модуля появляется один или несколько вариантов синтаксического дерева, которое представляет собой перечисление синтаксических отношений между лексемами. Так в предложениях выделяется подлежащее и сказуемое, определяются все дополнения, обстоятельства и определения, распознаются внутренние грамматические падежи. Этот процесс более универсален и в меньшей степени зависит от обрабатываемого естественного языка в том смысле, что если для построения синтаксического дерева необходимо знать грамматические правила языка, то на выходе уже получается более или менее универсальная структура.
5. Наконец, модуль семантического анализа, на вход которого подаётся список возможных синтаксических конструкций входной фразы (чаще всего состоящий из одной такой конструкции, но не всегда). Задачей семантического анализа является выявление «смысла» фразы, при этом во внимание должны приниматься не только смыслы каждой лексемы, входящей в неё, но и история коммуникации вместе с её контекстом. Семантический анализ является самой сложной задачей в рамках лингвистического анализа, и надо отметить, что до конца эта задача не решена именно в силу того, что пока мы не до конца знаем, как человек понимает смысл. Каким образом из сотен тысяч ассоциативных

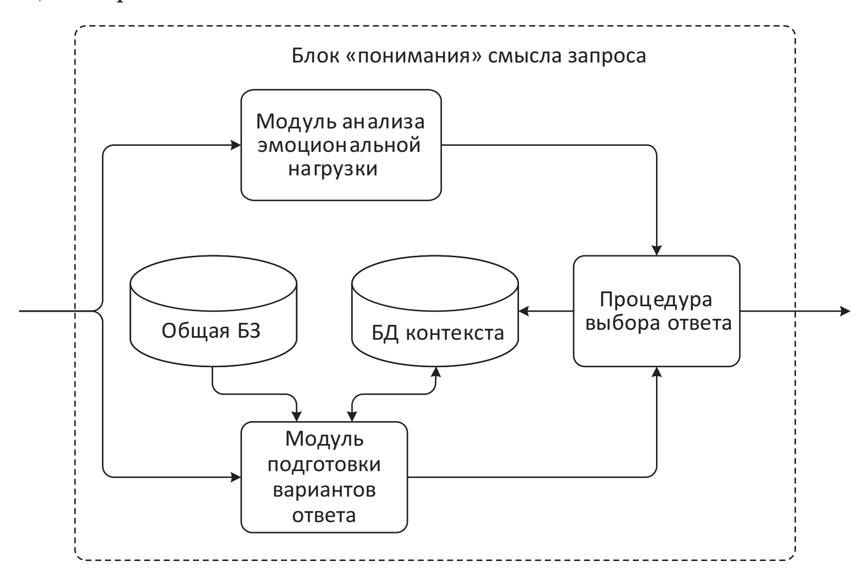
связей, которые задействуются в мозге во время коммуникативного акта, выбирается конкретный набор и как этот задействованный набор ассоциаций регулирует дальнейшее поведение человека – эти вопросы ещё ждут своего исследователя. Тем не менее возможно практическое применение уже сейчас имеющихся методов семантического анализа, в большей мере основанных на символьных вычислениях, – «понимание смысла» в данном случае видится как задействование правил или узлов семантической сети, которые наиболее релевантны входному высказыванию, и именно они определяют дальнейшее поведение интеллектуальной системы. Впрочем, восходящие методы искусственного интеллекта в этой задаче также могут помочь: если нейросетевую интеллектуальную систему обучать на входных фразах определённым классам поведенческих актов, то не будет удивительным то, что она начнёт адекватно реагировать на похожие входные фразы схожим поведением.

6. Также в составе лингвистического процессора имеются языковые базы данных. В первую очередь это база морфологии, при помощи которой осуществляется проверка на наличие ошибок, лексемный и морфологический анализы. Также есть базы для осуществления синтаксического и семантического видов анализа. В базе данных по морфологии приводятся начальные формы лексем и парадигмы их изменения. В базе данных по синтаксису описываются всевозможные синтаксические конструкции, свойственные языку. Наконец, в базе данных по семантике приводятся семантические отношения между лексемами языка, описываются классы синонимов и антонимов, даются определения терминам. Все эти данные крайне сильно зависят от того естественного языка, для анализа которого предназначается лингвистический процессор. Однако хорошая новость заключается в том, что сегодня большинство таких баз для самых распространённых естественных языков уже собрано и доступно разработчикам интеллектуальных систем.

Распознанный смысл входной фразы далее из лингвистического процессора передаётся в блок «понимания» смысла запроса. Этот компонент должен преобразовать смысл фразы в поведенческий акт системы, применив как общие знания об окружающей реальности, так и контекст коммуникаций. Кроме того, в послед-

нее время стала понятна важность распознавания эмоциональной нагрузки входной фразы, для того чтобы более качественно отвечать на неё с учётом «невербальных» сигналов пользователя.

В целом блок «понимания» смысла запроса выглядит следующим образом.




Подробнее разберём каждый элемент.

1. Модуль анализа эмоциональной нагрузки получает на вход результат семантического анализа входной естественно-языковой фразы и саму фразу, а на выходе возвращает описание эмоциональной нагрузки в виде значения из классификатора эмоций и, возможно, степени уверенности в этом значении. Эмоциональный анализ появился в рамках обработки естественного языка сравнительно недавно, но сегодня является неотъемлемой частью разговорных интерфейсов, поскольку выявление эмоций собеседника в диалоге позволяет выбрать стратегию ведения разговора. Сам анализ эмоциональной нагрузки осуществляется эвристически – по используемой лексике, по внеязыковым характеристикам фразы, по использованию «смайлов» и т. д. В целом эта область ещё менее изучена, нежели чистый лингвистический анализ.

2. Модуль подготовки вариантов ответов также получает на вход результат семантического анализа и на основании контекста беседы (хранится в базе данных контекста) и общих знаний о реальности и предметной области (хранятся в базе знаний системы) готовит один или несколько вариантов ответов. В качестве ответа может выступать или какая-либо естественно-языковая фраза (пока закодированная на внутреннем языке представления смысла), или последовательность действий системы, либо и то, и другое вместе.
3. Наконец, процедура выбора варианта ответа получает на вход перечень вариантов (может состоять из одного пункта) и результаты анализа эмоциональной нагрузки. В зависимости от определённых внутренних критериев, основанных на анализе прагматической составляющей обсуждаемой ситуации, эта процедура должна выбирать один вариант ответа, который возвращается в качестве итогового. Под прагматическим анализом здесь понимается то, что интеллектуальная система должна осуществлять выбор, принимая во внимание, как коммуникационный акт соотносится с ней, её текущим состоянием и целями. Также факт выбора конкретного ответа записывается в базу данных контекста беседы.

В целом так выглядит схема универсального алгоритма обработки естественного языка. К сожалению, сегодня пока ещё нет полноценных реализаций этой схемы, поскольку общая сложность естественного языка всё ещё недоступна для охвата программными средствами. Тем не менее разработки и дальнейшие исследования ведутся в следующих направлениях:

- символьная обработка на основе формальных грамматик;
- статистическая обработка;
- нейросетевая обработка с поиском и сопоставлением в пространствах подобия.

Во всех этих направлениях достигнуты определённые успехи, особенно в части анализа ограниченного подмножества языка с облегчённой грамматикой и ограниченным лексическим составом (так называемая «деловая проза»). Тем не менее остаётся много открытых вопросов и белых пятен. В частности, несмотря на то что теория формальных грамматик для полноценного символьного описания языка проработана очень детально, сами формальные грамматики для естественных языков не построены в полном объёме.

Статистическая обработка больших корпусов текстов не может считаться полноценным методом обработки естественного языка, так как в расчёт не принимается семантическая сторона высказываний. Вместе с тем результаты статистической обработки полезны как при построении формальных грамматик, так и при использовании искусственных нейронных сетей. Последние получили серьёзное применение в распознавании естественно-языковых запросов и генерации таких же ответов. Сегодня это одна из самых развивающихся областей искусственного интеллекта.

Обработка естественного языка традиционно используется для осуществления машинного перевода, для использования неформальных запросов к банкам данных и поиску в них, а также для организации простых интерфейсов между человеком и интеллектуальными системами. Последнее направление используется как для электронных помощников, так и в более узконаправленных решениях – экспертных системах и системах поддержки принятия решений. О них и поговорим в одном из следующих разделов…

## Раздел 3.3. Представление знаний

Теперь перейдём к представлению знаний, поскольку этот вопрос является одним из центральных как при построении систем слабого искусственного интеллекта, так и при исследовании природы сознания. Каким образом устроены знания человека? Как человеческий мозг их хранит? Каким образом знания в голове человека актуализируются? Эти и многие другие вопросы встают перед исследователями в полный рост, когда дело доходит до реализации. Особенно это касается чистого подхода, в рамках которого осуществляется символьная обработка знаний и вывод на её основе новых знаний.

Вопросы изучения знаний человека об окружающей реальности и абстрактных понятиях и методов познания действительности в рамках философии были выделены в отдельное направление эпистемологии. Далее в связи с развитием информационных технологий и возникновением исследований по искусственному интеллекту в рамках философии сознания была выделена когнитология – междисциплинарная область исследований, которая объединяет методы психологии и информатики для оптималь-

ного представления знаний определённой проблемной области и для решения конкретных задач в ней.

Чем же отличаются знания от данных? Ведь если рассматривать этот вопрос с точки зрения теории информации, то любое знание в конечном итоге можно закодировать при помощи последовательности битов, что также можно сделать и с данными, т. е. как будто бы знания и данные ничем не различаются. Это очень тонкий вопрос, и для его решения необходимо рассмотреть более конкретизированные свойства двух понятий, которые лежат вне прерогативы теории информации.

*Данные* – это представление фактов (информации) об окружающей реальности или абстрактных понятиях в форме, позволяющей осуществлять коммуникацию, интерпретацию и обработку информации человеком индивидуально или при помощи специальных средств. Это определение несколько затуманивает более наивное понимание того, что данные – это формализованные факты. Факты рассматриваются в рамках какой-либо *предметной области* и лежат в основе формирования гипотез и теорий. Таким образом, данные относятся только к фактам и используются в качестве базиса при формулировании ответов на конкретные вопросы о соответствующей предметной области. Использование данных для получения новой информации без использования механизмов логического вывода, суждений и умозаключений в общем понимании невозможно.

*Знания* – это совокупность утверждений об объективной реальности или абстрактных понятиях, свойствах объектов, закономерностях процессов и явлений (т. е. знания в своей основе схожи с данными), но в состав знаний входят правила логического вывода одних утверждений из других и правила использования утверждений для принятия решений. Главнейшее отличие знаний от данных состоит в том, что знания активны – появление информации о новых фактах или отношениях между ними может изменить способы принятия решений. Также, в отличие от данных, знания рассматриваются в рамках *проблемной области*, т. е. в совокупности предметной области и решаемых в ней задач.

Имеет смысл перечислить основные свойства знаний, которые отличают их от данных.

1. Знания имеют *более сложную структуру*, чем данные.
2. Знания задаются как *экстенсионально* (т. е. через набор конкретных фактов, соответствующих рассматриваемому

понятию), так и *интенсионально* (т. е. через свойства рассматриваемого понятия), а данные всегда задаются только экстенсионально.

3. Знания имеют *внутреннюю интерпретируемость*, т. е. возможность ассоциирования с элементом данных «избыточного» набора наименований тех множеств, в которые входит понятие.
4. Знания обладают *рекурсивной структурированностью*, т. е. возможностью расчленяться и объединяться в иерархические и родо-видовые отношения.
5. *Связность единиц знаний* определяет возможность установления различных отношений, отражающих семантику и прагматику связей отдельных явлений и фактов.
6. Наличие у знаний *семантического пространства с метрикой* позволяет определять близость и удалённость информационных единиц друг от друга.
7. *Активность знаний* определяет возможность формирования мотивов поведения, постановки целей, построения процедур их достижения.
8. *Функциональная целостность знаний* определяет возможность выбора желаемого результата, времени и средств получения результата, средств анализа достаточности полученного результата.

Наконец, прежде чем перейти к рассмотрению методов представления знаний, осталось рассмотреть пять уровней их организации. На первом уровне лежат знания в памяти человека – весь его основанный на эмпирическом опыте багаж знаний об окружающей реальности и правилах поведения в ней. На втором уровне лежат материализованные, но неформализованные знания – учебники, справочники, энциклопедии и т. д. На третьем уровне располагается так называемое «поле знаний», которое представляет собой полуформализованное, но уже структурированное описание знаний с первых двух уровней, т. е. из справочников и из головы экспертов. На четвёртом уровне поле знаний формализуется на каком-либо языке представления знаний в виде выбранной структуры. И наконец, на пятом уровне находятся знания в системе управления базами знаний.

Теперь рассмотрим отдельные формализмы для представления знаний. Не будем забегать вперёд и изучать то, как знания хранятся в искусственных нейронных сетях, а сосредоточимся

на символьных высокоуровневых формализмах. К таковым относятся:

- фреймы;
- продукции;
- семантические сети.

Рассмотрим каждый подробнее.

Термин «*фрейм*» ввёл в оборот Марвин Мински, когда разрабатывал системы распознавания сцен для навигации роботов. Первоначально под фреймом понималась минимальная единица описания наблюдаемой сцены, но потом понятие было расширено до описания объектов, явлений, событий, ситуаций, процессов и т. д. По своей сути иерархия фреймов представляет собой объектное описание проблемной области. Фреймы могут представлять описание абстрактной сущности (фреймы-шаблоны) и конкретных элементов объективной реальности, с которыми в текущий момент работает интеллектуальная система. У фреймов имеются слоты (как атрибуты у объектов в рамках объектно-ориентированного программирования), и эти слоты могут быть заполнены значениями, ссылаться на другие фреймы, в том числе и на фреймы-шаблоны, слоты которых, в свою очередь, должны быть заполнены при конкретизации понятия. При заполнении слотов могут вызываться сценарии для обработки воспринимаемых сигналов из внешней среды. В целом фреймы как способ представления знаний предлагают очень широкий набор методов работы со знаниями, но сам по себе этот формализм достаточно громоздкий. Сегодня существуют интеллектуальные системы, представление знаний в которых основано на фреймах, но их количество сравнительно мало с системами, основанными на других формализмах.

*Продукции* – это правила вида «ЕСЛИ… ТО…». Они соответствуют закону формальной логики ModusPonens и позволяют вывести некоторые заключения из имеющихся фактов. Сам формализм продукций предоставляет больше метаатрибутов для обработки таких правил, в том числе контекст продукции (проблемная область или её сегмент), условие применения продукции, различные коэффициенты значимости или достоверности, а также постусловие, которое может использоваться для изменения самой базы знаний при выполнении продукции. Для работы с продукционной базой знаний требуется машина вывода, которая получает на вход факты и при помощи продукций делает выводы. Ре-

зультаты такого вывода также попадают в список текущих фактов и, в свою очередь, могут использоваться для дальнейшего вывода. Машина вывода может осуществлять прямой вывод на основе уже упомянутого правила ModusPonens, и этот способ вывода используется для получения заключения на основе наблюдаемых факторов. Но также есть и обратная стратегия вывода, которая основана на правиле ModusTollens, и эта стратегия используется для получения возможных причин для текущих наблюдаемых проявлений (фактов). Некоторые более универсальные машины вывода одновременно могут использовать обе стратегии, смешивая их для достижения более сложных целей.

Наконец, *семантические сети* представляют собой направленные графы, в которых рёбра также несут на себе пометки. Обычно вершины представляют собой сущности проблемной области, а рёбра – некоторые отношения между ними. В общем смысле и вершины, и рёбра могут иметь произвольные пометки – их семантика определяется машиной вывода, которая обрабатывает семантическую сеть. Более того, для представления сложных ситуаций во многих проблемных областях желательно использовать расширение графов до отношений произвольной размерности. Это позволяют делать концептуальные графы, в рамках которых многомерные отношения представляются в виде отдельных узлов со специальными пометками. Вывод на семантических сетях осуществляется при помощи перехода от начальной вершины, в отношении которой надо осуществить вывод, ко всем вершинам, связанным с нею какими-либо отношениями, и далее до вершин, из которых уже нельзя совершить новых переходов. Результатом вывода являются все пути, семантика которых интерпретируется машиной вывода для конкретной цели и задачи.

Надо отметить, что чаще всего интеллектуальные системы одновременно используют несколько формализмов представления знаний. В частности, довольно часто объединяются семантические сети и наборы продукций. Это легко сделать потому, что у каждой продукции есть контекст её выполнения, и таким контекстом могут быть вершины семантической сети. Как только машина вывода доходит до какой-либо вершины в сети, на исполнение передаются все ассоциированные с этой вершиной продукции, и они исполняются в соответствии с принятой стратегией вывода и наличием означенных фактов из проблемной области. Это очень гибкий механизм, позволяющий представить

практически любые знания из произвольной проблемной области. С другой стороны, фреймы и продукции также можно объединить – в таком объединении фрейм будет представлять собой контекст для продукций, а сами продукции будут связаны со слотами. Как только управление переходит к какому-либо фрейму с необходимостью заполнить его слоты, продукции из слотов передаются в машину вывода на исполнение. Это позволяет также гибко реагировать на сигналы из окружающей среды – в продукциях, ассоциированных со слотами, могут быть записаны правила реагирования интеллектуальной системы на те или иные шаблонные и конкретизированные события в окружающей среде. Объединение семантических сетей и фреймов менее очевидно, но тоже возможно. Сами по себе фреймы уже объединены в сеть стандартными отношениями, выстраивающими иерархию фреймов. Если расширить набор отношений, то фактически получается семантическая сеть с фреймами в вершинах. Если при этом иметь возможность ассоциирования с вершинами (то есть фреймами) и их слотами наборов продукций, то получается наиболее общий и наиболее гибкий механизм представления символьных знаний.

Наконец, важным дополнением к описанным формализмам представления знаний будет *язык исполнения сценариев*. Это может быть язык в любой парадигме, но самая главная его особенность заключается в том, что в его рамках есть доступ к полной объектной модели проблемной области и он обладает достаточной функциональностью для её обработки. Это позволяет «на лету» создавать сценарии обработки знаний под нужды конкретных задач. Такие сценарии могут использоваться при заполнении слотов во фреймовой модели, при выполнении консеквента продукции или при активации вершины в семантической сети.

## Раздел 3.4. Машинное обучение

Следующим большим направлением в искусственном интеллекте является так называемое *машинное обучение*. Под этим термином понимается широкий набор методик, отличительной чертой которых является не прямой поиск решения, а постепенное обучение алгоритма способам нахождения решения. Это довольно общее описание, которое конкретизируется такими техниками и методами, как алгоритмы статистического поиска, методы оп-

тимизации и исследования операций, численные методы и многие эмпирические алгоритмы. Машинное обучение тесно связано с методами интеллектуального анализа данных и поиска в данных скрытых закономерностей, а также с методами извлечения и представления знаний. Тем не менее важным отличием является то, что часто машинное обучение может происходить самостоятельно без привлечения человека.

В этом разделе рассматривается исключительно индуктивное машинное обучение или обучение по прецедентам, когда на основе наблюдаемых данных интеллектуальный алгоритм делает обобщающие выводы, которые затем можно использовать на практике. Здесь, однако, уместно сделать лирическое отступление в область философии научного познания...

Индукция – это один из вариантов логического вывода, который основывается на переходе от частных положений к общему заключению. Индукция доказывает обобщённое умозаключение через перечисление частных случаев, при этом если для этого перечислены все имеющиеся частные случаи, составляющие всё возможное множество элементов общего случая, то индукция вполне правомерна и называется полной. Однако чаще в научном познании используется *неполная индукция*, когда на основе наблюдения за отдельными частными случаями строится обобщающая гипотеза. В этом случае наличие известных подтверждений гипотезы частными случаями не является её доказательством.

Проблема возникает тогда, когда на основе неполной индукции начинают делаться обобщающие выводы. Впервые заострил внимание на этой проблеме английский философ Томас Гоббс, а далее её развил и исследовал уже шотландский философ Дэвид Юм. Последний сформулировал два аспекта проблемы индукции.

1. Нет никаких оснований для обобщающих выводов относительно свойств класса объектов по факту наблюдений за отдельными экземплярами этого класса.
2. Нет никаких оснований для вывода предположений о будущем развитии событий на основании примеров того, как схожие события развивались в прошлом.

Важный вклад в изучение проблемы индукции сделали Иммануил Кант и Бертран Рассел, однако можно предположить, что наибольший интерес представляет мнение на этот счёт австрийского философа Карла Поппера, являющегося основоположником критического реализма, в рамках которого он сформулировал ключе-

вое понятие – *фальсифицируемость* (или *потенциальная опровержимость*). Сколько бы индуктивных наблюдений не производилось в поддержку той или иной научной гипотезы, она является лишь вероятной, но не абсолютно достоверной. При этом одного опровергающего наблюдения достаточно, чтобы отвергнуть рассматриваемую гипотезу. Это так называемая «гносеологическая асимметричность», вследствие которой К. Поппер выдвинул идею, что принцип верификации научных гипотез (осуществление подтверждающих проверок) должен быть заменён на принцип фальсификации – подтверждение истинности научных гипотез должно осуществляться не через подтверждение, а через опровержение.

Тем не менее естественный отбор поощрял развитие у древних людей механизмов, которые позволяли быстро построить правдоподобные гипотезы о причинно-следственных связях на основе неполной индукции. Древняя гоминида, которая убегала, заслышав рычание в джунглях, наверняка имела больший успех, чем та, которая начинала размышлять о возможных вариантах причин рычания, возможных вариантах их последствий и т. д. Всё это значит, что сегодня внутри головного мозга человека есть механизмы, которые делают обобщения на основе частичных наблюдений. Эти механизмы основаны на индуктивном обучении.

И более того, как говорил Бертран Рассел, неполная индукция в любом случае лежит в основе всякой науки, ибо шаг к познанию делается через наблюдение. А потому индуктивный подход и лёг в основу машинного обучения.

Машинное обучение, таким образом, относится к восходящему направлению искусственного интеллекта (к «грязной» парадигме), поскольку чаще пытается имитировать познавательную деятельность человека через эмуляцию низовых структур, чем через символьные вычисления и логики высших порядков. Ещё его называют «подходом, основанным на данных», поскольку упор делается не на заранее созданных алгоритмах, а на самостоятельной обработке больших массивов размеченных или даже неразмеченных данных, в которых осуществляется поиск закономерностей. Искусственные нейронные сети – типовой пример подхода, реализующего разнообразные методы машинного обучения. Тем не менее машинное обучение – это не только нейросети, так что в этом разделе кратко пройдёмся по алгоритмам и задачам, решаемым при помощи методов машинного обучения. Однако в любом случае тема эта настолько широка, что для детального

погружения в неё требуется использовать специализированную литературу и курсы.

Машинное обучение расширяет и обобщает классическую задачу аппроксимации, т. е. поиска достаточного приближения к наблюдаемым явлениям и объектам с целью дальнейшего моделирования и применения более «инвазивных» методов научного исследования в случаях, когда оригинальные объекты недоступны или к ним нельзя применять инвазивные методы воздействия. Суть задачи заключается в том, что интеллектуальная система должна получить на вход некоторое множество входных ситуаций и множество возможных ответных реакций на них. Объективно существует какая-то неизвестная зависимость между входами и выходами, и её необходимо восстановить, т. е. построить алгоритм, который для произвольной входной ситуации (даже не присутствовавшей в обучающей выборке) даёт достаточно точный ответ касательно реакции. В общем-то, это также классическая задача кибернетики по восстановлению содержимого «чёрного ящика» по его входам и выходам.

Классические задачи по аппроксимации функций и построению модели чёрного ящика обычно ставятся для числовых входов и выходов (чаще действительных, реже дискретных или векторных), а в реальных задачах машинного обучения ситуации могут быть произвольной природы. И, несмотря на то что объекты произвольной природы в любом случае можно закодировать дискретными кодами, общая задача машинного обучения часто появляется в условиях неопределённости, неточности, неполноты и других НЕ-факторов. Именно поэтому задача машинного обучения считается обобщённой.

Машинное обучение традиционно решает следующие задачи.

1. *Классификация*. Необходимо определить класс объекта на основе его признаков (например, определить то, что нарисовано на изображении, либо определить, является ли письмо спамом).
2. *Регрессия*. Определение (аппроксимация) и предсказание значений непрерывных параметров какого-либо объекта (например, предсказание того, как изменяется цена на акции).
3. *Кластеризация*. Группировка схожих объектов на основе их характеристик (например, разделение пользователей магазина на маркетинговые группы по их поведению).

4. *Ранжирование*. Выстраивание перечня некоторых объектов в порядке, максимизирующем некоторую функцию полезности (например, сортировка поисковой выдачи с учётом предпочтений пользователя).
5. *Снижение размерности*. Сжатие данных высокой размерности в низкую для визуализации или для использования в других алгоритмах искусственного интеллекта.
6. *Выдача рекомендаций*. Осуществление отбора ограниченного подмножества объектов из большого множества таким образом, чтобы элементы подмножества удовлетворяли заданным критериям (например, предоставление персонализированного выбора пользователям магазина для увеличения их лояльности).

Имеет смысл перечислить способы машинного обучения, каждый из которых либо применим в рамках конкретного метода, и тогда может использоваться, либо не применим, и тогда в рамках конкретного метода машинного обучения соответствующий способ не рассматривается. Итак, способов машинного обучения не так уж и много.

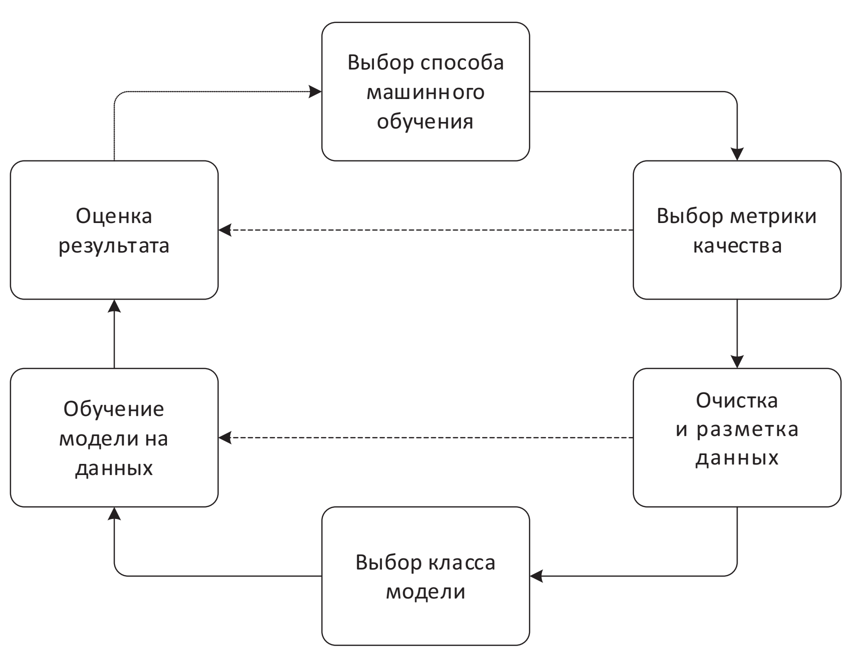
1. *Обучение с учителем*. Интеллектуальная система обучается на выборке пар «стимул – реакция», после чего может принимать на вход и варианты стимула, которых не было в обучающей выборке. Представляет собой ту самую классическую кибернетическую задачу с построением модели чёрного ящика. Наиболее известными примерами такого метода обучения являются метод опорных векторов (*supportvectormachine*) и отдельные классы искусственных нейронных сетей.
2. *Обучение с частичным привлечением учителя*. Разновидность обучения с учителем, в которой для тренировки используются как размеченные, так и неразмеченные данные.
3. *Трансдуктивное обучение*. Иначе называемый «обучением на прецедентах», этот способ не предполагает необходимости построить индуктивного обобщения на обучающей выборке размеченных данных, однако он строит трансдуктивные сети перехода от размеченных данных (прецедентов) к неразмеченным.
4. *Обучение с подкреплением*. Ещё один переходный способ, в котором роль «учителя» выполняет среда, из которой интеллектуальная система получает данные.

5. *Обучение без учителя*. Наиболее общий способ, в рамках которого интеллектуальная система пытается найти закономерности вообще самостоятельно без какого-либо воздействия «учителя», т. е. в условиях отсутствия разметки на данных. Может использоваться тогда, когда в данных предполагается какая-то структура, которую система может обнаружить и затем использовать для обучения.

Существуют ещё несколько намного реже используемых способов вроде «активного обучения» или «многозадачного обучения», но они могут быть выражены через перечисленные.

Для каждого способа машинного обучения имеются наиболее хорошо решаемые в рамках этого способа задачи. Так, для обучения с учителем подобными задачами являются классификация, регрессия и ранжирование. Обучение без учителя используется для кластеризации и снижения размерности. Обучение с подкреплением используется в случае интеллектуальных агентов, которые активно взаимодействуют со средой.

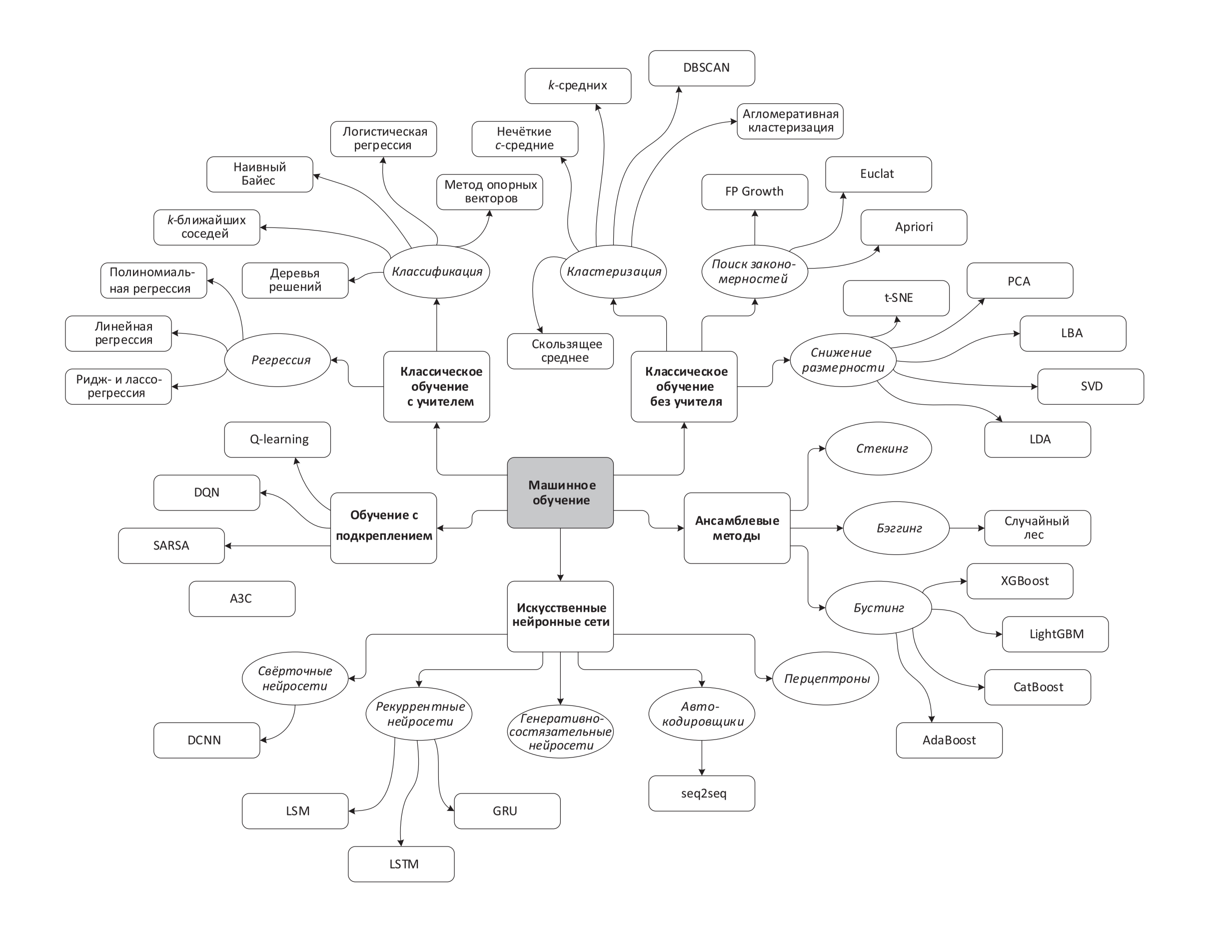
Общая схема решения задачи при помощи методов машинного обучения выглядит следующим образом.



Кратко опишем каждый шаг.

1. *Выбор способа машинного обучения*. Имеющуюся задачу необходимо привести к одной из типовых задач машинного обучения.
2. *Выбор метрики качества*. Необходимо выбрать метрику качества (критерий оценки) таким образом, чтобы можно было оценить как качество входных данных, так и качество полученного в результате работы модели машинного обучения результата.
3. *Очистка и разметка данных*. Входные данные для задачи должны быть очищены и размечены. Очистка предполагает удаление нерелевантной информации, нормализацию, структурирование и преобразование к входному формату. Разметка необходима, для того чтобы модель машинного обучения «понимала» семантику данных, с которыми ей предстоит работать.
4. *Выбор класса модели*. На этом этапе необходимо выбрать тот способ машинного обучения, который больше всего подходит для решения задачи: обучение с учителем для классификации, регрессии и ранжирования, обучение без учителя для кластеризации и уменьшения размерности, обучение с подкреплением для агентов, действующих в некоторой среде.
5. *Обучение модели на данных*. Прогон выбранной модели на очищенных и размеченных входных данных или их части для получения рабочей модели.
6. *Оценка результата*. Применение метрики качества на полученных результатах. На этом этапе часть входных данных может использоваться в качестве тестовых – на них обучение не проводилось, однако они используются для оценки работы обученной модели. В случае если результат устраивает, то работа завершается. Если результат неадекватен или не удовлетворяет критериям качества, либо полученный результат является входом для следующей модели, то цикл машинного обучения запускается заново.

Для решения типовых задач машинного обучения было создано большое количество математических методов и моделей, само перечисление и краткое описание которых заняло бы довольно много места. Однако здесь перечислим наиболее значимые методы и кратко охарактеризуем их.


Машинное обучение
Классическое обучение с учителем
Классическое обучение без учителя
Обучение с подкреплением
Ансамблевые методы
Искусственные нейронные сети
Регрессия
Классификация
Кластеризация
Поиск законо-мерностей
Снижение размерности
Полиномиаль-ная регрессия
Линейная регрессия
Ридж- и лассо-регрессия
Деревья решений
k-ближайших соседей
Наивный Байес
Логистическая регрессия
Метод опорных векторов
Нечёткие c-средние
k-средних
DBSCAN
Агломеративная кластеризация
Скользящее среднее
FP Growth
Euclat
Apriori
t-SNE
PCA
LBA
SVD
LDA
Q-learning
DQN
SARSA
A3C
Стекинг
Бэггинг
Случайный лес
Бустинг
XGBoost
LightGBM
CatBoost
AdaBoost
Свёрточные нейросети
DCNN
Рекуррентные нейросети
LSM
LSTM
GRU
Генеративно-состязательные нейросети
Авто-кодировщики
seq2seq
Перцептроны

На представленной диаграмме изображено то, что называется «Карта мира машинного обучения». Фактически всё машинное обучение можно разделить на пять больших разделов, каждый из которых можно дополнительно разделить по решаемым задачам или группам методов.

1. *Классическое обучение с учителем* – раздел машинного обучения, в котором ИИ-система использует чаще всего статистические методы для решения двух типовых задач машинного обучения: классификацию и регрессию. Классификация используется для размещения элементов дискретных множеств по заранее заданным классам, при этом тренировка системы происходит на множестве размеченных образцов. Регрессия – это практически то же самое, только про аппроксимацию и экстраполяцию непрерывных функций. С развитием численных методов и вычислительной техники все ранее разработанные математические и статистические модели получили «вторую жизнь» именно в рамках классического машинного обучения с учителем. Атрибут «классический» здесь (и в следующем наборе методов) обозначает, что в состав алгоритмов и методов не включаются нейронные сети, которые выделяются в отдельный класс методов.
2. *Классическое обучение без учителя* – другой раздел машинного обучения, в котором учитель, который размечает данные для классификации, отсутствует, а потому ИИ-система должна как-то научиться самостоятельно выделять те или иные признаки. Три главные задачи этого направления – кластеризация, поиск закономерностей и снижение размерности. Кластеризация – это та же классификация, только неизвестны названия классов, а системе предлагается самостоятельно разбить множество входных объектов на некоторые классы, выделив для этого какие-либо признаки. Иногда даже не даётся количество классов, но некоторые алгоритмы требуют этот параметр в качестве входного. Поиск закономерностей предполагает, что система обнаруживает во входных данных новые связи, которые можно использовать для решения тех или иных задач. Фактически это интеллектуальный анализ данных, который детально описан в одном из следующих разделов. Снижение размерности заключается в том, что иногда данные необходимо удобным образом представить для изучения человеком или другими

ИИ-системами, а для этого требуется исключить некоторое количество менее значимых факторов и свойств.

3. *Ансамблевые методы* – это три метода, которые, как ни странно, показывают очень мощные результаты даже по сравнению с нейросетевым подходом, но при этом просты до безобразия. Суть ансамблевых методов заключается в том, что для решения задачи запускается множество одинаковых или даже разных моделей, а потом результаты сравниваются, и среди них выбирается лучший. Стекинг как раз и основан на том, что входные данные обрабатываются несколькими разными алгоритмами, а их результаты передаются последнему, который и принимает решение по тому или иному критерию. Стекинг прекрасно параллелится, как и бэггинг, но на практике стекинг применяется редко, так как его точность по сравнению с двумя другими методами оставляет желать лучшего. Бэггинг берёт один алгоритм и многократно применяет его к случайным выборкам из входного множества данных, а в конце результаты работы усредняются. Наконец, бустинг является самым точным методом из числа ансамблевых, но он плохо параллелится. Его смысл в том, что обучение выбранных алгоритмов происходит последовательно, и каждый следующий алгоритм делает упор на те входные данные, на которых предыдущие алгоритмы отработали хуже всего. В качестве плюсов можно отметить точность классификации, а работает бустинг намного быстрее, чем нейронные сети.
4. *Обучение с подкреплением* – отдельное и довольно большое направление машинного обучения, в котором главной особенностью является то, что ИИ-система учится во взаимодействии со средой. В этом случае нет учителя, нет размеченных данных, а система совершает в своей среде функционирования некоторые действия. Среда же реагирует на действия и как-то «поощряет» или «наказывает» систему, в результате чего та учится. После проведения большого количества обучающих сессий ИИ-система начинает довольно адекватно вести себя в среде, в том числе и в ситуациях, в которых она не была во время обучения.
5. *Искусственные нейронные сети и глубокое обучение* – это огромный раздел машинного обучения, который сам по себе требует отдельного подхода к изучению. Здесь можно от-

метить, что чаще всего нейронные сети разделяют на пять больших классов архитектур: перцептроны, автокодировщики, рекуррентные сети, генеративно-состязательные сети и свёрточные сети. Все эти классы архитектур уже довольно подробно описаны в предыдущей главе.

Из всех перечисленных направлений машинного обучения осталось отдельно и кратко рассмотреть обучение с подкреплением, так как это направление, кроме методов машинного обучения, также частично захватывает эволюционные техники и теорию многоагентных систем. При этом идея подхода довольно проста.

Итак, пусть есть некоторая ИИ-система, которая функционирует в какой-то среде. У системы есть набор состояний, в которых она может находиться. Также у системы для каждого из её состояний есть набор возможных действий, которые она может осуществить. И для каждого действия в заданном состоянии среда возвращает некоторое число, воспринимаемое в качестве вознаграждения (если оно положительно) или штрафа (если оно отрицательно). Соответственно, целью ИИ-системы является переход из текущего состояния в целевое так, чтобы в процессе перехода максимизировать суммарное вознаграждение, получаемое от среды. Само собой разумеется, что от собираемого вознаграждения вычитаются получаемые штрафы.

Эта концепция гармонично ложится на то, как обучаются биологические системы. И с другой стороны, рассмотренные в одном из предыдущих разделов методы поиска также могут использоваться при решении поставленной задачи, так как поиск пути к целевому состоянию – это типичная задача поиска в пространстве состояний, в графе перехода между которыми на рёбрах написаны значения вознаграждения или штрафа от среды. Есть только один нюанс – у среды тоже может быть динамическое состояние, а потому вознаграждение или штраф является функцией от состояния среды. Поэтому статические методы поиска здесь не подходят. Кроме того, ИИ-система может предполагать только свои возможные действия для своего заданного состояния и состояния среды, которое она наблюдает. Больше никакой информации у системы нет. Именно эти ограничения не позволяют использовать тот набор методов поиска, который был рассмотрен, и для решения задачи требуется что-то принципиально новое. При этом задача прекрасно описывает поведение

робота в физическом мире или даже функционирование софтбота в мире виртуальном. Ведь не зря именно на методах обучения с подкреплением сегодня разрабатывают автономные автомобили и ИИ-решения, которые учатся играть в старые (и не очень) компьютерные игры (типа Mario Brothers).

Идея решения задачи обучения с подкреплением лежит в области динамического программирования. Чаще всего используется уравнение Беллмана и метод Монте-Карло для симуляции большого количества сессий взаимодействия системы и среды. Симуляция взаимодействия осуществляется до тех пор, пока не будет найдено приемлемое решение уравнения Беллмана для всех траекторий системы от текущего состояния к целевому. И такие симуляции – это как раз первый метод, который был назван Q-learning, где буква «Q» происходит от английского слова «quality» – качество. Но если среда слишком сложная, то можно поручить разбираться с ней нейронным сетям глубокого обучения, и тогда это метод DQN.

Заинтересованного читателя, которому хочется глубже погрузиться в методы машинного обучения, отправляю читать более узкую и специализированную литературу конкретно по машинному обучению, а здесь остаётся отметить, что наиболее интересными и развитыми являются такие модели, как искусственные нейронные сети и эволюционные алгоритмы, рассмотрение которых было проведено в предыдущей главе. Это важно ещё и потому, что именно эти две модели наиболее близки духу искусственного интеллекта.

## Раздел 3.5. Распознавание образов

Распознавание образов – это важная прикладная задача, часто сводящаяся к классификации и идентификации. Теоретические основы распознавания образов изучаются в рамках междисциплинарной теории распознавания образов, которая сегодня рассматривается как отдельное направление в рамках искусственного интеллекта. Практический аспект распознавания образов на текущий момент чаще всего сводится к использованию искусственных нейронных сетей, а в последнее время в связи с развитием технологий и повышением вычислительных мощностей – с нейронными сетями глубокого обучения.

Итак, распознавание образов – это в первую очередь математическая задача по определению класса заданного объекта. Математическая задача ставится довольно абстрактно, а потом эта абстракция «натягивается» на конкретные применения, и отсюда уже возникают задачи по распознаванию видеоизображений, аудиосигналов, каких-либо паттернов в последовательности значений измеряемых параметров и т. д. Если спускаться ещё ниже по лестнице абстракции, то появляются задачи по распознаванию следующих объектов: штрих-код, государственный регистрационный знак на автомобиле, лицо в толпе, инцидент на автомобильной дороге, загруженная мелодия или видеоролик, речь со всеми фонемами и наложенным шумом, факты подхода системы к аварийному режиму – и т. д. и т. п. На теоретическом уровне задача сводится к двум этапам – идентификации и классификации. Сначала образ выделяется из фона и идентифицируется, а потом он классифицируется, т. е. относится к одному из заданных классов.

Как должно быть понятно из небольшого набора перечисленных примеров, задача распознавания образов в большей мере касается именно визуальной информации. В других типах информации эта задача тоже может быть актуальной, но именно визуальная модальность наших органов чувств даёт человеку до 90 % информации об окружающем мире, поэтому именно изображения и видео как последовательность изображений стали главным объектом изучения теории распознавания образов. Вместе с тем можно отметить два подхода для решения этой задачи. Первый можно назвать «паттерн-матчингом», т. е. сопоставлением с образцом. Метод очень простой, но применим только тогда, когда информация, в которой производится распознавание образа, обладает простой структурой и небольшим объёмом. Ну, к примеру, классификация документов может быть выполнена при помощи сопоставления с образцами на основе деревьев решений.

Другое дело – визуальная информация. Она не только объёмна по своей природе, но и имеет сложную структуру в том смысле, что, кроме двумерной матрицы восприятия, которую надо соотнести с некоторым объектом, который может быть и трёхмерным, сам по себе образ объекта может быть повёрнут, иметь другой масштаб, иметь нестандартное или необычное цветовое решение, а также на него могут накладываться различные шумы и искажения. Например, если рассматривать распознавание символов, т.е., к примеру, букв, то для них может использоваться особый

шрифт, размер и различные декоративные элементы. Если распознавать трёхмерные объекты, например лица людей, то они могут быть повёрнуты не только в плоскости экрана, но и в любой другой плоскости или оси, что даёт произвольный ракурс – фас, анфас, профиль и т. д. Человеческий мозг прекрасно справляется с большинством таких искажений и преобразований, хотя и у него бывают сбои – существуют многочисленные оптические иллюзии. Но вот как сделать программу, которая будет распознавать образы?

Если обратиться к физиологии головного мозга, то сигналы от сетчатки глаз переходят на зрительную кору головного мозга через коммутирующее ядро в таламусе, которое выдаёт управляющие сигналы на мышцы век и зрачка. Сама же зрительная информация переходит в так называемую кору V1, т. е. первичную зрительную кору. Вот здесь и начинается самое интересное. Нейроны зрительной коры начинают возбуждаться в ответ на определённые примитивные паттерны, входящие в состав изображений. Например, огромное количество нейронов коры V1 возбуждается на контрастно яркую точку, находящуюся в тёмном окружении. Другие нейроны, наоборот, возбуждаются на тёмную точку в ярком окружении. Весь этот набор возбуждений передаётся далее на глубинную зрительную кору и зоны V2, V3, V4, V5, где распознаваемые образы всё больше и больше усложняются. Там, к примеру, есть нейроны, которые возбуждаются на контрастные линии, располагающиеся под различными углами. Другие нейроны возбуждаются на движение. Третьи – на определённые примитивные фигуры, четвёртые – на сплошную границу разделения контрастных сред. В конце концов, где-то в самой глубине есть нейроны, которые отвечают за узнавание лица человека, морды животного, автомобиля, здания и т. д., и уже эта информация передаётся куда-то совсем далеко в ассоциативно-теменную кору головного мозга, а из неё и в префронтальную кору, где, как предполагается, и находится то, что мы называем своим «Я». Отсюда же возбуждение распространяется по всей коре, задействуя все возможные ассоциативные связи, и таким образом складывается целостное квалиа восприятия того, что мы видим.

Фрэнк Розенблатт, как описано ранее, первым предложил новую революционную для того времени компьютерную модель, основанную на искусственный нейронах Мак-Каллока и Питтса, – это был перцептрон. Базовый перцептрон состоит из трёх слоёв, первый из которых называется сенсорным и получает

на вход визуальную информацию, закодированную тем или иным образом. Скрытый слой называется ассоциативным, и именно тут осуществляется создание «ассоциаций» между тем, что подаётся на вход, и выходным распознанным классом. Третий слой, выходной, активирует тот или иной нейрон в зависимости от того, к какому классу принадлежит распознанный объект. Модель довольно простая, но тем не менее она оказалась очень мощной. Можно сказать, что практически любую задачу распознавания образов можно решить при помощи перцептрона, если повысить вычислительные мощности. Кроме того, были разработаны расширения базовой модели – несколько вариантов многослойных перцептронов, использование других вариантов искусственных нейронов, использование обратных связей между слоями. Всё это позволяет сейчас говорить о том, что перцептрон и основанные на нём методы составляют вторую методику распознавания образов в дополнение к сопоставлению с образцами.

Но технологии развиваются, так что появилась новая архитектура искусственной нейронной сети, которая получила название «свёрточная». Свёрточная сеть очень эффективно решает задачу распознавания образов. Автор архитектуры, Ян Лекун, взял некоторые идеи из устройства нервной системы млекопитающих и, в частности, зрительной коры головного мозга. Структура свёрточной нейронной сети более или менее детально описана во второй главе. Обычно работа такой сети описывается как постепенный переход от конкретных особенностей распознаваемого образа к абстрактным деталям всё выше и выше по иерархии уровней абстракции. Например, в применении к изображениям это обозначает, что первые слои обнаруживают в частях изображения различные типовые элементы вроде разнонаправленных отрезков, а на последних слоях нейроны уже активируются в ответ на наличие в составе изображения каких-либо форм, конкретных объектов или образов. Суть операции свёртки, из-за которой сеть получила своё название, заключается в том, что каждый фрагмент входного образа поэлементно умножается на матрицу свёртки, результат этого перемножения суммируется, а итоговая сумма передаётся в аналогичную позицию выходного образа для заданного слоя. И получается постепенное сворачивание изображения во всё меньшие и меньшие по размеру слои. Так что на самых глубоких слоях можно найти нейроны, которые реагируют на наличие в исходном изображении, например, лица

человека, при этом не важно, как лицо повёрнуто. Это в какой-то мере повторяет работу зрительный коры в неокортексе млекопитающих.

При этом не стоит считать, что искусственные нейронные сети являются единственным методом распознавания образов. В математике разработан ряд методик, которые позволяют кластеризовать входные данные, а затем определить класс объекта чисто аналитическими методами, часто не требующими использования формализма нейросетей. Впрочем, последние также могут реализовывать такие аналитические методы. В частности, метод опорных векторов для кластеризации вполне успешно реализуется при помощи нейронной сети с соответствующей архитектурой. В общем виде методы распознавания образов можно классифицировать следующим образом:

- сопоставление с образцами (функции близости);
- дискриминантный анализ;
- статистические методы;
- искусственные нейронные сети;
- эвристические методы.

Для использования функций близости распознаваемые образы обычно представляются в виде векторов в пространствах высокой размерности. Каждое измерение такого пространства соответствует какому-либо признаку распознаваемого объекта. Признаки для представления объектов должны выбираться таким образом, чтобы максимально дифференцировать объекты и их классы. Таким образом, метод сопоставления с образцами представляет собой вычисление функции близости для распознаваемого объекта с некоторыми объектами-эталонами, являющимися типовыми представителями своих классов. И для распознаваемого объекта в качестве класса выбирается тот, для которого функция близости по отношению к объекту-эталону принимает минимальное значение. Таким образом, это метод классификации.

Функция близости может быть произвольной. Она определяет метрику пространства, не более. Обычно в качестве такой функции используется квадратный корень из суммы квадратов разниц между значениями по каждой координате для распознаваемого образа и эталона, и тогда такая функция определяет «обычное» расстояние между объектами, как в евклидовом пространстве. Но могут быть и иные метрики, которые подбираются специально под задачу.

Дискриминантные функции используются для кластеризации. Если есть пространство признаков объектов распознавания, в котором такие объекты так или иначе распределены, то при помощи дискриминантных функций можно «нарезать» пространство на кластеры, после чего любой распознаваемый объект будет попадать в тот или иной кластер. Наиболее известным методом дискриминантных функций является метод *k*-средних.

Статистические методы, как и функция близости, решают задачу классификации и делают это при помощи формулы Байеса для оценки апостериорной вероятности вхождения распознаваемого объекта в тот или иной класс.

Но сегодня, после того как искусственные нейронные сети глубинного обучения превзошли все иные методы индуктивного машинного обучения, в том числе и в решении задачи распознавания образов, говорить о других методах можно лишь в контексте их историчности.

## Раздел 3.6. Интеллектуальный анализ данных

Следующей прикладной задачей является интеллектуальный анализ данных, который представляет собой набор методов для обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Здесь каждое слово в этом определении важно. Давайте посмотрим на эти понятия.

Предположим, что есть некоторый набор данных. Допустим, что имеется огромный массив наблюдений за погодой, скажем, в Москве за весь XX век – каждый день в разных точках Москвы снимались и записывались показатели погоды: температура окружающей среды, атмосферное давление, влажность воздуха, геомагнитная обстановка, фаза Луны и т. д. Можно записать сотни параметров. И это будут те самые «большие данные», про которые сегодня так любят все говорить.

При помощи интеллектуального анализа этих больших данных можно было бы получить большое количество новых знаний – *скрытых закономерностей*, которые могут быть связаны как с географическим положением и действовать только для Москвы, так и общи-

ми для всей планеты. Каждая закономерность может выражаться при помощи продукционного правила. Например: «Если в первой половине ноября пошёл снег, то зима будет тёплой». Напоминает что-нибудь? Конечно, это форма так называемых погодных примет. Наши предки всю жизнь занимались интеллектуальным анализом данных, только делали это бессознательно. И конечно же, такие правила могут иметь различную степень достоверности и корректности.

Указанное правило про снег в первой половине ноября должно быть ранее неизвестно. Зачем искать в данных закономерности, которые уже известны? А это значит, что ИИ-системы, которые осуществляют дата-майнинг, – это ещё одно название для интеллектуального анализа данных, – должны иметь базу знаний по той предметной области, для которой анализ осуществляется. Поскольку нельзя просто так взять и статистическими методами обработать все записанные факторы – результаты надо сравнивать с тем, что уже имеется. И тут пригодятся методы верификации знаний – иногда могут получаться противоречивые знания, для которых необходимо противоречия разрешать.

Следующий аспект – выявленные в процессе дата-майнинга знания должны быть нетривиальны. Это объяснить просто – тривиальные знания никому не нужны. Что интересного в правиле «Если идёт снег, то температура окружающей среды ниже нуля»? Это будет правило с высшей степенью достоверности, но оно не несёт никакой полезной нагрузки. Поэтому ИИ-система также должна уметь оценивать правила на тривиальность и общий «здравый смысл», если можно так выразиться.

Следующий очень важный аспект – это то, что выявляемые знания должны быть практически полезными. Опять же неприменимые на практике знания бесполезны и никому не нужны, кроме как, возможно, архивариусам, которые поместят их на полку и внесут в пыльный каталог. Ну, к примеру, если в корпусе погодных данных обнаружится закономерность типа «Раз в сто лет в Москве в июле бывает снег на протяжении двух-трёх дней, но спрогнозировать его появление можно только за неделю до выпадения». Такое знание с практической точки зрения бесполезно. Мало того, что явление очень редкое, его ещё и заранее спрогнозировать нельзя.

И наконец, четвёртый аспект – это доступность интерпретации. Под интерпретацией понимается объяснение, толкование, раскрытие внутреннего смысла знания. Если знание невозможно проинтерпретировать, то на его основе невозможно построить

объяснение принятому решению, хотя все предыдущие аспекты выявленного знания могут иметься в наличии – неинтерпретируемое знание может быть новым, нетривиальным и практически полезным. Но невозможность интерпретации сводит это всё на нет.

В качестве примера можно привести что-то очень сложное, невероятно запутанное, типа: «Если каждый второй вторник всех чётных месяцев лета на протяжении последних пяти лет шёл дождь, а также через сорок два дня после последнего дождя был заморозок, но при этом через пять дней после него не светило солнце, а вечером того же дня был туман, а на следующее утро роса, то либо пятого января следующего года будет оттепель, либо вторая неделя апреля послеследующего года будет нетипично жаркой». На самом деле это уже интерпретация, так как правило было облечено в словесную форму. Обычно неинтерпретируемые знания очень сложно завернуть в подобные словесные оболочки, но выглядят они примерно так.

Хорошо. Вот пусть имеется корпус данных, и теперь понятно, какими свойствами должны обладать выявляемые из них знания. Что делать дальше? Тут необходимо разъяснить, что обычно дата-майнинг пока осуществляется в автоматизированном режиме, т. е. при участии человека, а не в автоматическом, когда участие человека не требуется. Этими вопросами, т. е. дата-майнингом, занимается дата-саентист. Дата-саентист должен последовательно осуществить следующие операции:

- осуществить подготовку данных;
- выбрать модель для анализа данных;
- подобрать параметры для выбранной модели;
- обучить модель;
- проанализировать качество обученной модели;
- проанализировать сами данные;
- проанализировать выявленные закономерности.

Имеет смысл кратко рассмотреть все эти операции.

Перед запуском процесса интеллектуального анализа данных эти самые данные необходимо подготовить. Какой бы «умной» ни была бы модель анализа, всё равно данные необходимо очищать, формализовывать и размечать. Если опять вернуться к примеру с корпусом погодных данных, то эти данные должны быть размечены так, чтобы ИИ-система понимала, где температура, где давление, а где просто комментарий, который не нужно принимать во внимание во время анализа. Кажется, что эта процедура

тривиальна, однако на самом деле очень часто крах происходит из-за того, что на вход системе для анализа поданы некачественные данные. Ведь принцип GIGO, т. е. «мусор на входе – мусор на выходе» никто не отменял. И надо отметить, что естественный интеллект работает точно так же: на мусорных данных будут сделаны неадекватные выводы. Можно привести забавный пример из области биоинформатики – данные о генах попадали в электронную таблицу, которая «умно» конвертировала названия генов в другой тип данных, и результат анализа был совершенно неадекватным. И вот эта типовая ситуация с электронными таблицами случается постоянно. Даже прямо сейчас кто-то этим занимается.

Затем для анализа данных выбирается модель. Сегодня есть просто огромное количество общих и довольно конкретизированных моделей, так что дата-саентисту надо в них ориентироваться. И вряд ли сегодня ему придётся самостоятельно реализовывать какую-либо модель анализа. На выбор есть большой набор статистических, эвристических, нейросетевых моделей и других моделей машинного обучения. Во всём этом зоопарке необходимо разбираться.

У каждой модели дата-майнинга обычно есть какие-либо параметры. Наверное, только самые простые модели не имеют параметров, а просто работают с данными. У более или менее продвинутых моделей параметры есть. Например, это могут быть такие параметры, как уровень разброса значений, который можно игнорировать, т. е. принимать все значения из определённого интервала, как одно и то же значение. Или в нейросетевых моделях есть большое количество разных параметров, связанных с типами нейронов, количеством слоёв и т. д. И обычно нет определённых требований и правил настройки – дата-саентист часто действует на основе своего опыта и интуиции.

Следующий шаг – обучение модели. Для этого из всего корпуса подготовленных данных выбирается обучающая выборка, которая, в свою очередь, делится на собственно обучающую и тестовую. Модель с настроенными параметрами прогоняется на обучающей выборке, а потом при помощи тестовой выборки осуществляется следующий шаг работы – анализ качества обученной модели. Может так получиться, что модель очень сильно настроилась именно на данные обучающей выборки и с другими уже не работает. Это называется «переобученностью». В таком случае делается откат на два шага назад и снова начинается подбор параметров модели. А возможно, и откат ещё на шаг назад для выбора новой модели.

Если качество обучения модели устраивает, то можно нацелить её на весь корпус данных. Результатом является набор выявленных знаний – тех самых скрытых закономерностей. Дата-саентист оценивает адекватность этих знаний и использует для этого те самые четыре аспекта, которые были упомянуты ранее: знания должны быть ранее неизвестными, нетривиальными, практически полезными и доступными для интерпретации. Все или хотя бы большинство выявленных знаний должны строго ложиться в эти аспекты, чтобы результаты можно было назвать адекватными. Если это не так, то опять происходит откат назад к выбору модели, подбору её параметров и обучению.

Как видно, работа дата-саентиста пока ещё больше похожа на какую-то магию. В целом это действительно так, поэтому по сегодняшним временам труд дата-саентистов оплачивается очень хорошо. Особенно тех, кто собаку съел на анализе сложных и неструктурированных данных. Но также ведутся работы по переводу этого процесса в полностью автоматический режим. И как ни странно, помогут в этом именно технологии искусственного интеллекта, а именно эволюционный подход.

## Раздел 3.7. Обработка НЕ-факторов

В этом разделе приведён очень краткий обзор современных подходов, методов и технологий получения, представления и обработки знаний с такими НЕ-факторами, как нечёткость, неопределённость, неточность и недоопределённость. Под получением понимаются механизмы (возможно, разработанные только на теоретическом уровне) перевода информации, полученной от эксперта, из структурированных текстов или из баз данных, в некоторые формализованные структуры на каком-либо языке представления знаний. Под представлением понимается форматирование формализованных структур, полученных на этапе извлечения, с целью долгосрочного хранения (например, в базах знаний). И наконец, под обработкой понимается не только машинный вывод на знаниях, полученных на этапе извлечения, но также и верификация этих знаний, т. е. уже формализованных структур.

В области когнитологии, входящей в состав нисходящей парадигмы искусственного интеллекта, разработано достаточно большое число методов для работы с экспертами, а также предложены

различные способы обработки полученных в результате взаимодействия с ними результатов. Кроме того, создан целый ряд программных средств, автоматизирующих процессы извлечения знаний из экспертов, специальных текстов на естественном или структурированном языке и баз данных.

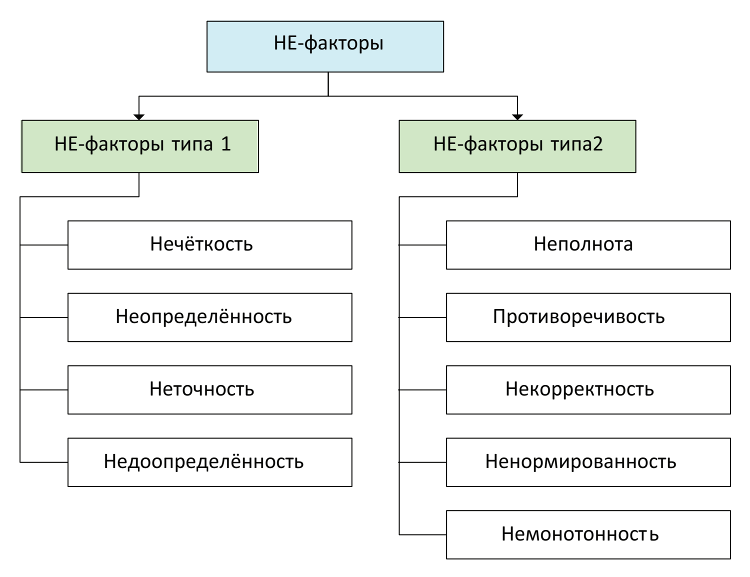
Но при этом малоисследованными всё ещё остаются вопросы приобретения знаний при формировании непротиворечивых баз знаний с упомянутыми НЕ-факторами. Эти вопросы имеют большое значение, поскольку знания, извлечённые из экспертов, как правило, содержат различные виды НЕ-факторов, в связи с чем соответствующие методы и процедуры приобретения знаний должны обеспечивать возможность извлечения и обработки не полностью известной (недостоверной) информации. В целом решением проблем, связанных с представлением и обработкой каждого из НЕ-факторов, занимаются самостоятельные направления исследований, где для этих целей создаются специальные математические аппараты и формализмы, наиболее известными из которых являются аппарат нечётких множеств, методы недоопределённых моделей и так называемое программирование в ограничениях, а также теории возможностей и неопределённости.

Имеет смысл более формально рассмотреть, что же представляют собой процессы обработки знаний.

- *Приобретение знаний* – процесс получения знаний от эксперта или из каких-либо других источников и формализация этих знаний для последующего использования их в системах, основанных на знаниях. Приобретение знаний является одним из видов процесса получения знаний, при этом собственно под приобретением знаний понимается получение знаний из источников знаний при помощи использования программных средств поддержки деятельности инженера по знаниям и эксперта.
- *Представление знаний* – процесс, реализующий ответы на два вопроса: «*Что представлять?*» и «*Как представлять?*». Первый вопрос – это вопрос определения состава знаний, и его важность определяется тем, что решение именно этой задачи обеспечивает адекватное отображение моделируемой проблемной области. Второй вопрос, в свою очередь, разделяется на две в значительной степени независимые задачи: как организовывать (структурировать) знания и как представить знания в выбранном формализме.

- *Обработка знаний* – вывод на знаниях, а также логическая проверка знаний, т. е. верификация, хотя зачастую под последним процессом понимается также и синтаксическая и семантическая проверка полей и баз знаний.
- *Верификация* – процесс логической проверки полей и баз знаний. В случае использования продукционных систем процесс верификации знаний сводится к проверке и, возможно, устранению или исправлению определённых правил или наборов правил продукционной системы, выбранных по критериям верификации. К таким критериям относятся наличие определённых НЕ-факторов в правилах (противоречивые и неполные правила), логические ошибки в правилах (транзитивно-замкнутые правила), излишество правил и т. д.

НЕ-фактором называется некоторое понятие, которое лексически, синтаксически и семантически отрицает какое-либо свойство или аспект знания, как, например, противоречивость (отрицает непротиворечивость знания), неточность (отрицает точность знания) и т. д. НЕ-факторы можно классифицировать следующим образом.




Обобщённая классификация НЕ-факторов

Приведённая классификация разбивает всё множество НЕ-факторов на два основных класса – *НЕ-факторы типа 1* и *НЕ-факто-*

*ры типа 2*. Это разбиение проведено с точки зрения возможности приобретения знаний с НЕ-факторами в автоматизированном режиме из эксперта. Так, НЕ-факторы типа 1 можно извлекать из эксперта в автоматизированном режиме при помощи определённых эвристических механизмов. С другой стороны, НЕ-факторы типа 2 не подлежат извлечению из источников знаний первого рода (экспертов), так как они вообще должны быть по возможности устранены из систем, основанных на знаниях. Для устранения проявлений НЕ-факторов типа 2 также используются различные технологии, например для обнаружения неполноты можно использовать механизмы интеллектуального анализа данных.

Важно рассмотреть главные НЕ-факторы, которые наиболее часто встречаются в извлекаемых знаниях.

- *Нечёткость*. Аппарат нечёткой логики был разработан в середине XX века Л. А. Заде как расширение аристотелевой логики. Нечёткая логика является бесконечнозначной логикой, оперирующей значениями истинности из интервала [0; 1], причём 0 – это полная ложь, 1 – полная истина, а все промежуточные значения интерпретируются в зависимости от решаемой задачи.
- *Неопределённость*. Под неопределённостью понимается случай, когда к значениям некоторых параметров проблемной области эксперт приписывает некоторую степень уверенности (которая, в свою очередь, может быть сложной природы). Например, эксперт может явно описывать свою уверенность в высказываниях о проблемной области в виде числа из интервала [0; 1] либо оценивать уверенность подынтервалом из этого же интервала.
- *Неточность* – это один из наиболее часто встречающихся НЕ-факторов, так как он проявляется в знаниях тогда, когда при извлечении оцениваются некоторые параметры, полученные при помощи измерительных приборов, которые имеют свою погрешность измерения. Именно погрешность измерения обусловливает то, что измеренные параметры неточны.
- *Недоопределённость* – это частичное отсутствие знаний о значении какого-либо параметра (измеримого или нет). В случае измеримых параметров недоопределённость и неточность можно легко приводить друг к другу, однако существует чёткое разграничение. В случае недоопределённости

частичное отсутствие знаний можно восполнять, постепенно доопределяя параметр, а неточные измеренные параметры самодостаточны сами по себе, так как зачастую повышать точность измерения для решения конкретной задачи не имеет смысла (например, бессмысленна точность в один метр при решении гипотетической задачи о перемещении планет в околосолнечном пространстве).

Фактически для понимания того, как НЕ-факторы обрабатываются в интеллектуальных системах, имеет смысл привести таблицу для выделенных НЕ-факторов в разрезе процессов обработки. Следующая таблица сводит информацию о методах обработки четырёх НЕ-факторов.

| НЕ-фактор | Приобретение | Представление | Обработка |
|---|---|---|---|
| Нечёткость | Существует большое количество прямых и косвенных методов для получения нечётких знаний из экспертов и их групп, а также извлечения нечёткости из письменных источников на основе обработки естественного языка | Лингвистические переменные и функции принадлежности нечётких переменных, входящих в их состав, – нечёткие числа LR-типа, кусочно-линейные функции принадлежности | Прямой и обратный нечёткий вывод – имеется, как минимум, 6 различных стратегий нечёткого вывода, в том числе и авторский максиминный метод на кусочно-линейных функциях принадлежности. Кроме того, используется нечёткая арифметика |
| Неопределённость | Прямые методы с запросами степени неопределённости, а также лингвистические методы и метод репертуарных решёток | Точечные и интервальные степени неопределённости | Метод Байеса (в том числе вероятностные сети), схема Пиэрла и метод Демпстера–Шейфера |
| Неточность | Прямые методы с запросом интервалов неточности и лингвистические методы | Интервал неточности или мера погрешности измерений | Интервальная арифметика и её обобщения на нестандартные операции. Сравнение интервальных величин |
| Недоопределённость | Прямые методы с запросом интервалов недоопределённости и лингвистические методы | Доверительные интервалы | Интервальная арифметика и её обобщения на нестандартные операции. Сравнение интервальных величин |

Таблица фактически является индексом для поиска описания упомянутых методов в специализированной литературе. Формат этой книги не позволяет дать детальное описание каждого метода.

Видно, что к настоящему времени наиболее полно проработана математическая теория, предоставляющая механизмы для всесторонней обработки нечёткости. Эта ситуация вполне правомерна, так как сам этот НЕ-фактор был введён в рассмотрение ещё в 50-х годах XX века. С того времени было создано несколько сотен тысяч научных публикаций, в основном приходящихся на японских и европейских учёных. Эти публикации раскрывают все возможные нюансы нечёткой математики во всех её аспектах.

Менее разработанной оказывается теория обработки неопределённости, хотя и относительно этого НЕ-фактора написано довольно большое количество публикаций. Это связано с тем, что многие исследователи считают неопределённость одним из видов нечёткости, что позволяет в какой-то мере применять и те аппараты, которые разрабатывались для работы с нечёткостью. Кроме того, неопределённость можно обрабатывать различными вероятностными и статистическими методами (что показывают соответствующие методы обработки, полученные из расширения классических вероятностных методов), поэтому здесь также можно учитывать огромное количество разработок в области теории вероятности и математической статистики.

Самыми малоисследованными НЕ-факторами из выделенных оказываются неточность и недоопределённость. И если математические теории для работы с неточностью более или менее разрабатывались, так как интервальная арифметика необходима во многих прикладных областях человеческой деятельности, то недоопределённость как отдельный НЕ-фактор была выделена совсем недавно, что дополнительно сказывается на количестве публикаций на тему работы с недоопределённостью.

## Раздел 3.8. Экспертные системы и системы поддержки принятия решений

Развитие экспертных систем началось чуть ли не с началом эры искусственного интеллекта, и разработка этого класса систем была вторым существенным прорывом в исследованиях. Экс-

пертные системы основаны на трёх ранее рассмотренных технологиях – символьных вычислениях, обработке естественного языка и представлении знаний. Когда идея экспертных систем только родилась у пионеров искусственного интеллекта, казалось, что достаточно посадить эксперта, который тщательно формализует свою проблемную область на выбранном формализме представления знаний, и искусственная интеллектуальная система сможет самостоятельно вырабатывать решения в любых ситуациях этой проблемной области. Однако этот наивный подход столкнулся с серьёзными проблемами самого разного характера.

В первую очередь возникли проблемы организационного характера, если так можно выразиться. Хотя в целом они не являются непреодолимыми, разработка экспертных систем серьёзно «провисла». Дело в том, что эксперты не стремятся делиться своим знанием, особенно в целях его неограниченного распространения и применения. Ведь если экспертная система будет решать все задачи за человека в заданной проблемной области, то зачем нужен эксперт? Другими словами, это вполне обоснованное опасение специалиста в том, что если он поделится своим знанием, то тем самым «размоет» свою полезность и востребованность. Тем не менее вопрос вполне решаем с учётом того, что большая часть действительно нужных знаний изложена в письменных источниках – статьях и книгах. И их можно использовать для построения баз знаний интеллектуальных систем.

Тем не менее в конечном итоге даже нашлись эксперты, которые были крайне заинтересованы в создании подобных систем. И они начали работать над базами данных. И тут-то обнаружился следующий проблемный аспект – экспертные знания, в общем-то, не так просто формализовать. Их прямой перевод на язык формальной логики возможен только в очень ограниченных случаях, а в большей части экспертные знания представляют собой переплетение тех самых «НЕ-факторов» – они неполны, часто противоречивы (особенно если рассматривать знания нескольких экспертов), неточны в силу неточности измерений и фундаментальной неопределённости реального мира. Часто знания выражаются на нечётком естественном языке, и для решения повседневных задач этого достаточно, но совершенно недостаточно для формализации в виде баз знаний. Даже появление профессии когнитолога или инженера по знаниям не помогло – пришлось развивать математический аппарат. В итоге

и эта проблема была побеждена, и сегодня существует большое количество формализмов и методов обработки знаний с НЕ-факторами, как это было показано в предыдущем разделе.

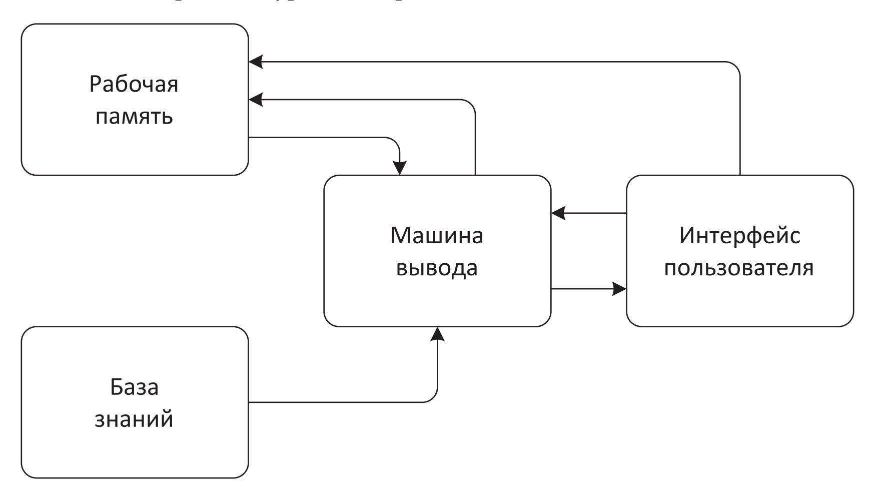
Но и наличие НЕ-факторов в знаниях экспертов не стало такой серьёзной проблемой, как динамика в знаниях. Знания – это не статическая, заранее и навсегда установленная догма, которую можно запрограммировать один раз и дальше пользоваться всегда. В процессе исследований оказалось, что знания – это эволюционирующая сущность, истинная природа которой пока всё ещё скрывается от исследователей. А потому было очень сложно формализовать то, как знания изменяются в процессе своей эволюции. Понятно было, что этот процесс можно описать такими словами, как «обучение», «актуализация» и т. п., однако от этого проще не становилось. Как осуществлять обучение экспертной системы? Когда наступает тот самый момент, характеризующий необходимость актуализировать знания о проблемной области? Кто должен актуализировать знания? Может ли система самостоятельно актуализировать и дополнять свою собственную базу знаний? Это немногие из широкого ряда вопросов, которые встали перед первыми разработчиками экспертных систем. На некоторые из таких вопросов полноценных и исчерпывающих ответов нет до сих пор.

Хотя в целом описанные сложности привели к приостановке развития экспертных систем и методологии их построения, до сих пор есть как отдельные экспертные системы в очень узких проблемных областях, так и целые довольно проработанные фреймворки и оболочки для построения экспертных систем. Что характерно, постепенно этот инструментарий впитывал в себя все те новые математические формализмы, которые позволяли справиться с НЕ-факторами и динамичностью знаний. И сегодня существующие, и развивающиеся фреймворки для экспертных систем представляют собой очень продвинутое программное обеспечение. Хотя надо признать, что в основном все они основаны на обработке символьной информации и исповедуют нисходящую парадигму искусственного интеллекта.

В итоге надо отметить, что именно проблемы и сложно решаемые вопросы при построении экспертных систем сделали основной вклад в явление, которое потом назвали второй зимой искусственного интеллекта. Постепенно весь ажиотаж и смелые надежды в духе «сейчас как опишем все знания всех экспертов, и будет система самой умной» сошли на нет, и многие направ-

ления исследований и проекты в этой области были свёрнуты. Но, конечно же, отдельные энтузиасты продолжили развитие методологии инженерии знаний, так что сегодня эта область постепенно возрождается.

Однако перед тем как перейти к подробному рассмотрению новых методик и технологий, которые пришли на смену экспертным системам, необходимо чуть более детально посмотреть, что же находится «под капотом» у традиционных экспертных систем. Это будет полезно для понимания того, как должен быть устроен символьный искусственный интеллект (или подсистема верхнего уровня в гибридном искусственном интеллекте). Итак, вот общая архитектура экспертной системы.




Кратко опишем компоненты экспертных систем, представленных на этой диаграмме.

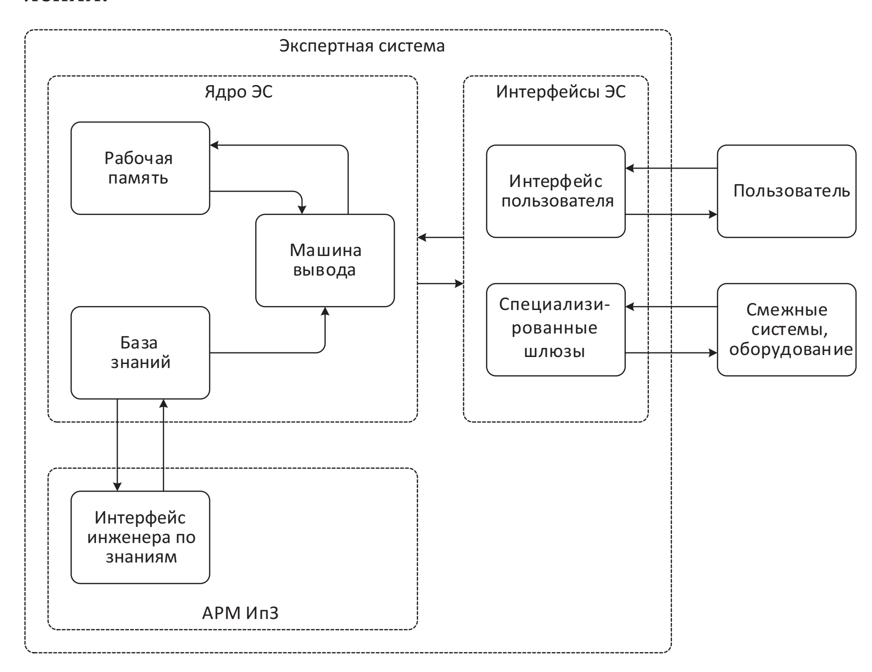
1. *База знаний* содержит знания экспертной системы о проблемной области. В режиме штатной работы обычно база знаний не изменяется, а потому информация из неё берётся машиной вывода в режиме «только чтение». Изменение базы знаний осуществляется в режиме обучения системы экспертом. Однако в последнее время наибольший интерес представляют экспертные системы, которые имеют возможности самообучения в процессе работы с пользователями.
2. *Рабочая память* содержит факты, которые вводятся пользователем при помощи интерфейса (либо как-то иначе добы-

ваются экспертной системой из внешнего мира, например с сенсоров). При помощи фактов машина вывода активирует знания из базы знаний и выводит новые факты, которые также помещаются в рабочую память.

3. *Машина вывода* – это ядро экспертной системы. Она получает знания из базы знаний и факты из рабочей памяти, а также запросы пользователя через интерфейс взаимодействия с ним. Далее происходит цикличная работа с рабочей памятью и, при необходимости, с пользователем, когда на основе имеющихся фактов и знаний осуществляется вывод новых фактов, их уточнение у пользователя или запрос у него дополнительных фактов. В конечном итоге машина вывода получает результаты вывода и выдаёт их пользователю через интерфейс с ним.
4. Наконец, *интерфейс пользователя* представляет собой тот механизм, при помощи которого пользователь взаимодействует с экспертной системой. В самом начале развития этого направления планировалось, что интерфейс пользователя будет естественно-языковым, однако с анализом естественного языка возникли существенные проблемы, а потому чаще всего использовался формализованный ввод (иногда даже похожий на естественно-языковой). Сегодня тема естественно-языкового ввода возрождается в связи с развитием методов анализа естественного языка и появлением большого количества мессенджеров, имеющих возможность использования чат-ботов в качестве универсального метода языкового доступа к системам.

Экспертная система может быть интегрированной в информационную систему более высокого уровня, тогда, как уже было указано выше, в её рабочую память информация может поступать не только посредством интерфейса пользователя, но и из других систем или, например, с сенсоров, если экспертная система используется при управлении каким-нибудь технологическим процессом. С другой стороны, интегрированная экспертная система может выдавать свои заключения не только конечному пользователю, но и другим системам или даже исполнительным устройствам. Другими словами, сегодня экспертные системы могут использоваться как полноценные и законченные блоки для создания комплексных систем управления с возможностью взаимодействия не только со специально обученными пользователями, но и с другими системами.

Это позволяет нарисовать более целостную диаграмму того, что представляет собой экспертная система в своём окружении. С учётом вышесказанного получается полноценная обобщённая кибернетическая система, взаимодействующая со средой. И такое взаимодействие осуществляется в парадигме символьных вычислений.

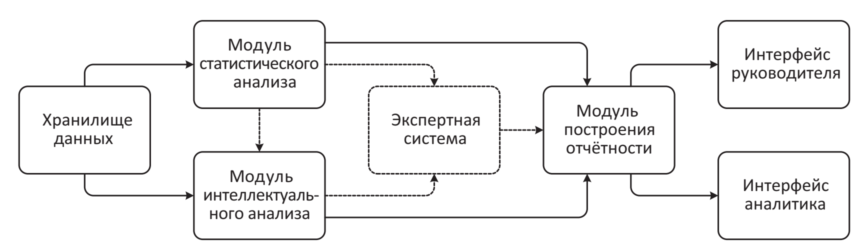



Интересным является вопрос: в составе каких более крупных систем могут включаться экспертные системы? Самый тривиальный ответ: «В составе систем автоматического управления и автоматизированных систем управления». Это традиционные области использования экспертных систем, особенно в хорошо формализуемых проблемных областях. Например, системы автоматического управления технологическими процессами вполне могут управляться при помощи достаточно небольших экспертных систем, базы знаний которых содержат несколько сотен продукций.

Но самым главным классом систем, в которых используются экспертные системы в качестве элементов, являются системы поддержки принятия решений. С развитием аналитических методов и появлением новых математических формализмов для обработ-

ки больших объёмов данных, для поиска закономерностей, в том числе скрытых, системы поддержки принятия решений получили второе дыхание, и сегодня этот класс систем используется в большом количестве областей применений для снятия с аналитиков и руководителей рутинной работы по сбору и анализу больших массивов информации и подготовке к принятию решений. Так что внимательно посмотрим на такие системы – какие методы в них используются, какие составные элементы в них имеются, каковы взаимосвязи между элементами и всё остальное.

Общая структура систем поддержки принятия решений выглядит следующим образом.




Кратко рассмотрим все элементы, приведённые на схеме, за исключением экспертных систем, которые были детально описаны ранее.

1. *Хранилище данных* – первичный элемент системы, в который собираются все данные об объекте управления или проблемной ситуации, относительно которых система осуществляет поддержку принятия решений. Как и в случае с экспертной системой, в хранилище попадают данные как от пользователей, так и с различного рода датчиков, устройств и иных систем. В отличие от базы данных, в хранилище осуществляется первичная обработка данных таким образом, чтобы потом их можно было эффективно использовать для анализа и визуализации. Происходит очистка и преобразование данных, а их хранение осуществляется в специальном формате, который позволяет быстро манипулировать ими. Всё это помогает двум другим компонентам системы – модулям статистического и интеллектуального анализа – быстро обрабатывать огромные массивы информации.
2. *Модуль статистического анализа* предназначается для применения различных методов математической статистики

и смежных дисциплин для анализа больших объёмов информации (статистических выборок), в частности поиска корреляций и причинно-следственных связей, анализа трендов, определения статистических закономерностей и т. д. Большинство методов статистического анализа давно разработано и апробировано, однако временами появляются новые, а потому важно, чтобы этот элемент системы был открытым для использования новых методов.

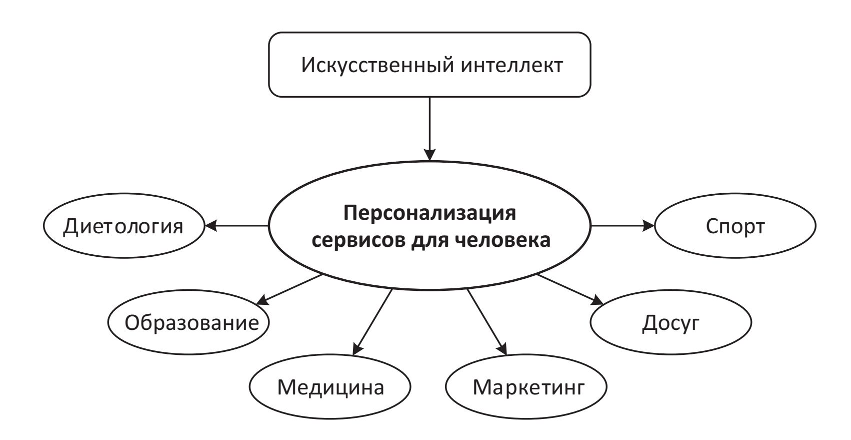
3. *Модуль интеллектуального анализа* используется для проведения глубокого анализа больших объёмов данных методами, которые позволяют обнаруживать скрытые закономерности, нетривиальные и практически полезные интерпретации информации в данных. В первую очередь это методы из разряда технологий DataMining. Этот модуль сам по себе реализует многие техники искусственного интеллекта, начиная с генетических алгоритмов и искусственных нейронных сетей и заканчивая отдельными методами символьных вычислений. При этом в рамках интеллектуального анализа данных также могут проводиться и отдельные ветви статистического анализа, либо результаты такого анализа могут браться из соответствующего модуля.
4. *Модуль построения отчётности* – фактически центральный модуль, преобразующий результаты анализа и (или) выводы экспертной системы в вид, доступный лицу, принимающему решения. Основная задача этого модуля – доступная и эффективная визуализация найденных закономерностей и знаний для руководителя, который на основании этой информации может принять взвешенное и оптимальное решение. В этом и заключается сама сущность поддержки принятия решений.
5. *Интерфейс аналитика*. В процессе принятия решений и его поддержки важна роль аналитика, который готовит окончательный вариант решения или варианты альтернатив на основе того, что рекомендует система. Аналитик модифицирует процедуры извлечения и обработки информации из неструктурированных больших объёмов данных, он же проверяет окончательные результаты работы системы и её объяснения того, как она получила эти результаты. Поэтому для аналитика имеется интерфейс, при помощи которого он

может взаимодействовать с системой. Он более функционален, нежели интерфейс руководителя.

6. *Интерфейс руководителя*. Лицо, принимающее решение, является главным пользователем систем поддержки принятия решений, поскольку именно для него и готовятся варианты решения с объяснениями, так чтобы этот ответственный руководитель выбрал итоговое решение и своей подписью придал ему официальный статус. Для этих целей в составе системы имеется специальный интерфейс с упрощённым набором функций, который позволяет руководителю получить варианты решений, увидеть их обоснования и либо принять один из вариантов, либо отправить все варианты на доработку, снабдив их дополнительными комментариями.

Сегодня системы поддержки принятия решений используются во многих областях жизни. В первую очередь необходимо отметить, что они являются неотъемлемой частью деятельности различных штабов, действующих в условиях крайне ограниченного времени – при ликвидации последствий чрезвычайных ситуаций, при очень быстро развивающихся событиях внешнего характера, при невозможности задействования людей при выполнении каких-либо работ. В этих случаях системы поддержки принятия решений используются в рамках так называемого ситуационного управления, когда подобная система должна за крайне ограниченное время обработать большие объёмы динамической информации о развитии ситуации, над которой осуществляется управление, и выдавать лицам, принимающим решения, обоснованные прогнозы, планы и решения.

Но не только государственные службы, задействованные в предупреждении и ликвидации последствий чрезвычайных ситуаций, используют системы поддержки принятия решений. Многие коммерческие компании пользуются системами этого класса для развития своего бизнеса. Это относится к организациям, работающим в области телекоммуникационных технологий, розничной продажи, обслуживания массового клиента, банковской сферы. Системы поддержки принятия решений в этих областях находят различные скрытые закономерности в поведении клиентов, что позволяет организациям тонко настраивать тарифную политику или даже организовывать персонализированный маркетинг.




В будущем прогнозируется использование таких систем не только в связке с людьми, принимающими решения, но и автономно, когда искусственные интеллектуальные системы самостоятельно принимают решение. Это позволит повысить скорость принятия решений, исключить человеческий фактор в критических областях, а также персонализировать подход к каждому конкретному человеку. И это ожидается не только в части маркетинга (персонализированные рекламные акции и продажи), но и в таких областях, как образование, здравоохранение, юриспруденция, и многих других. Хотя, конечно же, остаются вопросы этического плана и ответственности за решения, ожидается, что эти вопросы будут сняты, а уровень решений будет настолько качественным, что говорить об ответственности интеллектуальной системы не придётся.

В этом случае имеет смысл говорить о постепенной эволюции от систем поддержки принятия решений к интеллектуальным системам принятия решений.

## Раздел 3.9. Робототехника

Робототехника – это одна из прикладных дисциплин в рамках искусственного интеллекта, которая находится на стыке искусственного интеллекта и других прикладных технологий – электроники и электротехники, механики и мехатроники, кибернетики и информатики. Более того, в последнее время специалисты в области робототехники активно изучают психологию и физиологию

восприятия и дальнейшей обработки сенсорных сигналов в нервной системе живых организмов.

Возможно, это будет звучать несколько необычно, но робот – это всего лишь техническая кибернетическая машина, действующая в нашей реальности. Это самое общее определение, которое можно дать данному термину. Очень конкретные примеры роботов – это автономный автомобиль, робот-пылесос или робот-манипулятор на производственной линии какого-либо промышленного производства.

Конечно, робот – это не только и не столько антропоморфное создание типа робота R2D2 из саги «Звёздные войны». Большинство роботов, которые уже работают на Земле, – это совершенно не похожие на человека устройства, часто выполняющие очень узкие функции. Например, на заводах-автоматах целые производственные линии состоят из технологических роботов-манипуляторов, которые производят различные элементы конечной продукции и далее собирают, проверяют и выпускают её.

Поскольку это книга по искусственному интеллекту, а не по робототехнике, то здесь будут рассматриваться только автономные роботы с элементами интеллектуальных функций. Это довольно сильно ограничивает область рассмотрения, и здесь речь пойдет лишь о таких кибернетических машинах, которые работают в нашей реальности и имеют интеллектуальную систему управления. Что такое «кибернетическая машина»? Это тройка (сенсоры – система управления – исполнительные устройства). Робот в этом случае представляет собой как бы тело для ИИ-системы, при этом у него есть инструменты как для функционирования в нашем мире, так и для подключения к миру виртуальному для получения из него бездонного количества информации.

Можно кратко рассмотреть примеры проблемных областей, в которых могут работать интеллектуальные роботы. Во-первых, это, конечно же, промышленность. На смену роботизированным технологическим линиям, работающим по заданным алгоритмам, могут прийти интеллектуальные роботы с самообучением, которые будут самостоятельно перестраивать технологическую линию для выполнения новых задач. Такие линии уже появляются, а в совокупности с технологией 3D-печати и другими аддитивными технологиями они смогут быть очень гибкими.

Во-вторых, это бытовые роботы для выполнения рутинных работ по домашнему хозяйству. И это будет очень замечательно, когда

не только уборка в квартире будет выполняться роботом-пылесосом, но и вообще все бытовые работы. Ну вот не нравится кому-нибудь стирать и гладить рубашки по утрам, так пусть же это делают роботы.

В-третьих, это вопросы обеспечения безопасности. Помните огромных боевых человекоподобных роботов? Вот это оно. На самом деле тут ожидается развитие в сторону миниатюризации автономных летающих дронов с возможностью роевого интеллекта, о чём будет немного рассказано в следующем разделе.

В-четвёртых, это медицина – уже сегодня роботы-хирурги достигли очень высокой точности, и в ближайшем будущем они превзойдут людей, если уже не превзошли.

В-пятых, это роботы в области сервиса – младший медицинский персонал, учителя, страховые агенты, чиновники, клерки и все-все-все, чья деятельность связана с физическими действиями и может быть легко автоматизирована и интеллектуализирована. Представляете, заходит кто-нибудь в страховую компанию застраховать свой автономный левитирующий автомобиль, а там его встречает робот, который только выносит пакет готовых документов на подпись. Впрочем, зачем куда-то идти и что-то печатать не бумаге? Не будет этого, конечно же.

Есть ещё одна важнейшая область применения роботов. Представьте себе окружение, которое очень вредно или опасно для человека. А робот может там находиться, даже если у него начнут «плавиться мозги». Что-то совсем опасное, типа рабочей зоны ядерного реактора. Или же что-нибудь недоступное, типа дна Марианской впадины или кровеносных сосудов человека. Роботы уже используются для посещения таких мест и работы там.

И вот тут мы приходим к понятию «аватара». Штука, показанная в одноимённом фильме, – это, конечно, пока фантастика, но исследования в этом направлении ведутся. Гвоздь технологии – прямой нейроинтерфейс от человека к роботу. Если подсоединить робота напрямую к человеческому мозгу, то роль системы управления будет выполнять оператор со своим естественным интеллектом. Тогда такого робота, который и называется аватаром, можно отправлять в любые места, куда он может попасть в силу своих физических особенностей, но куда не может попасть человек. И человек сможет управлять им, получая обратную связь из того же интерфейса. Здесь больше нерешённых проблем, чем каких-то открытий, но направление очень бурно развивается. Тем более что оно имеет крайне прикладное значение – нейро-

протезы для конечностей и вообще любых органов будут очень востребованы.

Главная проблема с аватарами – скорость прямой и обратной связи. Мы не сможем отправить аватара на Марс и управлять им там, лёжа в кресле на Земле, поскольку сигнал дойдёт до Красной планеты только минут через десять в зависимости от того, как располагаются Земля и Марс друг относительно друга. За двадцать минут путешествия сигнала туда-сюда аватара там уже могут украсть марсиане, либо он свалится в какую-нибудь канаву, которых на Марсе много. Но вот управление аватарами с орбиты – вполне реальное дело. И возможно, именно при помощи этой технологии мы и начнём колонизировать Солнечную систему и добывать на астероидах полезные ископаемые.

Что ж, это было некоторое лирическое отступление. Теперь давайте посмотрим, какие роботы выделяются в рамках робототехники. Итак, раз робот – это «тело» для ИИ-системы, которая должна через него взаимодействовать с физической реальностью, можно классифицировать роботов по способу их передвижения в нашем мире. Здесь выделяются два больших класса, и второй класс подразделяется на несколько подклассов.

- *Манипуляторы* – стационарные или иногда передвижные автоматы, которые состоят из устройства для взаимодействия с объектами окружающего пространства (что-то типа «руки») и системы управления им. Обычно у руки такого робота имеется большое число степеней свободы, которые позволяют ей ловко управляться с тем, на что она воздействует. Такие роботы-манипуляторы предназначаются для выполнения рутинных операций на конвейерах или же для работы в опасных средах.
- *Мобильные роботы* – машины, которые могут передвигаться и при этом также выполнять различные иные действия. Передвижение мобильных роботов может осуществляться при помощи:
  - *колёс*, и тогда это колёсные роботы, которые могут ездить по горизонтальным и наклонным поверхностям, самостоятельно доставляя себя в требуемые точки пространства;
  - *гусениц*, и тогда это гусеничные роботы, которые могут выполнять такие же функции, как и колёсные, только перемещаться могут на более агрессивных и сложных поверхностях;

— *педипуляторов*, т. е. шагающих устройств, которых может быть от двух (как у человека) до многих. Роботы с педипуляторами фактически могут перемещаться там же, где может ходить человек на своих ногах;
— *других средств*, и здесь можно выделить ползающие, плавающие и летающие робототехнические системы, предназначенные для работы в соответствующих средах. Этот подкласс мобильных роботов наиболее узок, хотя в последнее время он получает серьёзное развитие в связи с большим количеством исследований в области беспилотных летающих аппаратов.

Также можно классифицировать роботов по способу управления ими. В этом случае выделяются следующие классы роботов.

- *Биотехническое управление*. Управление роботом осуществляется непосредственно человеком через прямую связь. В качестве интерфейса управления используются панели с кнопками и рычагами, программные интерфейсы, а в последнее время – и нейроинтерфейсы. Фактически это те самые аватары, о которых шла речь ранее.
- *Автоматизированное управление*. Человек-оператор задаёт программу функционирования, которую исполняет робототехническая система, но при этом у оператора есть возможность вмешаться в её исполнение практически на любом шаге.
- *Автоматическое управление*. В этом случае человек исполняет лишь функцию целеполагания, а сам робот самостоятельно решает поставленную целевую задачу.

Сегодня роботы в основном используются в промышленности и для решения специфических задач в области обеспечения безопасности. Тем не менее всё больше и больше применений достижения робототехники находят в области гражданского сервиса – роботы-игрушки, бытовые роботы для выполнения рутинных задач, роботы для взаимодействия с клиентами и даже секс-роботы. Большое внимание уделяется исследованиям в области взаимодействия с людьми и эмпатии, для того чтобы создать роботов-сиделок для престарелых и инвалидов.

Имеет смысл более детально изучить возможности использования технологий искусственного интеллекта в промышленности. Сегодня под общим термином «промышленность» понимается вообще любое производство материальных ценностей на основе некоторой цепочки технологических переделов. То есть добы-

ча полезных ископаемых, сельское хозяйство, металлургические предприятия, пищевые предприятия и т. д. – это всё промышленность в контексте данного рассмотрения. Само собой разумеется, что промышленность лежит в фундаменте всякой экономической системы и, как следствие, почти всякого государства. Конечно, на Земле уже имеются примеры государств, в которых промышленность полностью отсутствовала бы, но это в основном карликовые или совсем неразвитые государства, зарабатывающие туризмом и почтовыми марками. В общем, промышленность – это один из столпов экономики. Рассмотрим же теперь, что могут дать технологии искусственного интеллекта для развития этой области.

В вопросах управления производством можно руководствоваться такими же принципами, которые используются при управлении государством. Ведь замечено, что при переходе количества задействованных на предприятии сотрудников в своём числе через какой-то порог само предприятие начинает напоминать «маленькое государство», имеющее внутреннюю и внешнюю политику, множество органов и институтов «власти» и «администрирования», которые в непосредственном производстве продукции не участвуют. И тут открывается широчайший простор для автоматизации и интеллектуализации технологических и деловых процессов – от оптимизации количества сотрудников, выполняющих рутинные операции, до внедрения системы поддержки принятия решений на самом высшем уровне управления предприятием.

Намного больший интерес представляет собой рассмотрение промышленных предприятий в виде кибернетических машин. Ведь кибернетика – это довольно универсальная наука и способ познания объективной реальности и процессов в ней. И экономическая кибернетика является сравнительно развитым направлением кибернетической мысли. Так что рассмотреть субъект экономических отношений в виде кибернетической машины вполне возможно. Она получает на вход сырьё, инструменты и производственную силу, производит внутри себя последовательность технологических переделов и выпускает на выход продукцию. Получается «чёрный ящик» с набором входов и выходов. Теперь такие чёрные ящики можно соединять друг с другом в кибернетические схемы – на вход одного предприятия подаётся продукция, выпущенная другим предприятием.

Попытки построить такую схему для всей экономики государства или хотя бы небольшого региона, не говоря уже о мировой

экономике, обычно приводили к фиаско. Пока ещё неизвестны успешные результаты таких работ, даже если снизить требования и попытаться просто соединить предприятия и отрасли на одной схеме без учёта различных временных и пространственных ограничений. Тем не менее эта задача вполне по силам интеллектуальной системе, которая могла бы не только разработать такую кибернетическую схему, но и потом при помощи неё осуществлять оптимизацию всей экономической системы в целом, организовывая бережливое производство и поставки точно в срок. Возможно, это советская плановая утопия. Не буду отрицать, что какие-то аналогии провести можно. И собственно, если у Госплана СССР не получилось обустроить экономику, то это не значит, что метод плохой и надо полагаться только на «невидимую руку рынка». Прогнозирование, планирование и эффективная реализация с обратными связями – всё это может быть вполне реализовано при помощи методов искусственного интеллекта.

Теперь можно рассмотреть примеры, как технологии искусственного интеллекта уже могут помогать в промышленности. Первая технология – это так называемый «интернет вещей», который, конечно, часто относят к отдельному направлению техники. Тем не менее если рассматривать эту сущность под углом зрения данной книги, то можно увидеть две несомненные технологии искусственного интеллекта – анализ больших данных и построение многоагентных систем различных конфигураций. Интернет вещей является одним из наиболее популярных направлений современного технологического развития, и при организации одноранговых или иерархических сетей между агентами, работающими на производстве или в сельском хозяйстве, это направление очень востребовано. Особенно в сельском хозяйстве – там вообще огромное количество возможностей для применения. В итоге получается простая штука – огромное количество сенсоров собирают большие данные, которые анализируются системой управления на базе деревьев решений, нейронных сетей или гибридных решений, после чего осуществляется воздействие на технологические циклы производства или непосредственно на сырьё, материалы, инструменты.

Ещё, например, есть большое количество автономных решений для выполнения каких-либо специальных функций там, где человеку быть опасно или попросту невозможно. Здесь помогает робототехника и создание различных специализированных устройств. Робототехника помогает и при развёртывании гибких

производственных линий. Ну а оптимизация бизнес-процессов с учётом возможностей по замене персонала решениями на базе искусственного интеллекта или повышения эффективности работы персонала по тем же причинам открывает широкие возможности по внедрению искусственных интеллектуальных систем в бэк-офисе любой компании.

Наконец, можно было бы отметить и такую возможность. Уже упоминалось о системах поддержки принятия решений на высшем стратегическом уровне управления технической системой. В своё время отдельными коллективами очень серьёзно и глубоко прорабатывались вопросы комплексной автоматизации промышленных предприятий, при которой все технологические, производственные и деловые процессы замыкались в единый контур автоматизации и посредством интеграционной платформы объединялись в одну комплексную систему управления. Данные в этой системе курсировали от места их первичного зарождения по всем местам, где они обогащались и использовались, достигая как раз той самой СППР на самом верху, в которой вся информация о производстве и всех аспектах его деятельности появлялась точно в срок без запаздывания, что позволяло бы принимать взвешенные решения на основе актуальных данных. Тогда проработанные решения далеко опередили своё время, но сегодня они становятся всё более востребованными, поскольку технологии шагнули вперёд. И опять же появился интернет вещей и всё остальное. Так что сегодня системы комплексной автоматизации для предприятий любого уровня сложности вполне возможны.

## Раздел 3.10. Многоагентные системы и роевой интеллект

Уже было упомянуто, что многоагентные системы – это одно из наиболее динамично развивающихся направлений искусственного интеллекта. Вместе с тем это также и наименее проработанное направление, так как своё развитие эта теория начала примерно с 90-х годов XX века. Поэтому в этом разделе приводятся лишь основы того, что есть в теории многоагентных систем.

Начнём с азов. Многоагентная система – это система, образованная из нескольких взаимодействующих друг с другом автономных интеллектуальных агентов. Тут важно именно то, что

агентов должно быть несколько, т. е. хотя бы два, и они должны взаимодействовать друг с другом.

Идём дальше… Интеллектуальный, или рациональный, агент – это некоторая сущность, которая функционирует в среде, получает от неё и от других агентов информацию, воздействует на среду и передаёт другим агентам информацию. Самое главное то, что агенты действуют рационально, т. е. пытаются достичь своей цели, затратив на это минимально возможное количество ресурсов с учётом всех ограничений, накладываемых средой и самой многоагентной системой.

Агенты должны быть автономными, т. е. иметь определённую свободу действий и принятия решений. Конечно, тут опять надо вспомнить про ограничения среды и всей многоагентной системы. Но вопрос именно в децентрализации – если система централизованная, то говорить о многоагентности сложно. Впрочем, на уровне всей системы могут быть поставлены общие цели, но потом каждый агент действует самостоятельно.

Теперь переходим к взаимодействию самих агентов. Когда какой-либо агент пытается получить то, что ему непосредственно недоступно, то он может начать просить либо близких к нему агентов, либо всех агентов системы о помощи. Это будет просьба о предоставлении того, что ему нужно, либо просьба выполнить какое-либо действие, либо ещё какая-то просьба.

Что происходит далее? Агенты вокруг спрашивающего оценивают его просьбу на предмет способности её выполнить, а потом те из них, кто готов выполнить просьбу, отправляют агенту стоимость выполнения, выражаемую в тех самых ресурсах, минимизацию затрат которых надо обеспечить. Изначальный агент выбирает исполнителя с учётом стоимости и оценки качества выполнения. Если консенсус достигнут, первый агент передаёт ресурсы, второй выполняет, остальные получают отказ.

Вот здесь можно остановиться и подумать, какие примеры агентов и многоагентных систем могут существовать? Ну, в животном мире всё просто – социальные насекомые, стаи птиц и косяки рыб, стадные животные, человек. В мире информационных технологий всё немного сложнее. Разработать программных агентов несложно, так чтобы они взаимодействовали в какой-либо виртуальной среде.

А что, если программно-технические агенты действуют в нашей реальности? Тогда люди и различные объекты инфраструктуры тоже входят в состав среды, в которой они действуют и с которой

взаимодействуют. А быть может, люди сами могут быть включены в состав многоагентной системы? Тут открывается очень много возможностей. В качестве примеров автономных агентов в этом случае можно привести промышленных роботов, пассажирский транспорт, автономные автомобили, боевые дроны.

Давайте рассмотрим некоторые варианты классификации агентов и многоагентных систем. Это позволит более точно определять параметры создаваемых ИИ-систем для решения тех или иных задач. Рассмотрим несколько различных шкал классификации как агентов, так и многоагентных систем, которые в большинстве случаев являются ортогональными, что позволяет охарактеризовать эти объекты с различных точек зрения.

Начнём со шкал классификации агентов. Первая очевидная шкала – это шкала «естественный – искусственный». Многоагентные системы могут включать в свой состав большое количество разнообразных агентов, и они, естественно, совсем не обязаны быть одного типа. Вполне возможно, что многоагентная система состоит как из искусственных, так и из естественных агентов, к которым относятся любые живые существа, которые могут так или иначе взаимодействовать с окружающей средой и другими агентами, – насекомые, животные, люди, организации и прочие типы естественных агентов.

Вторая шкала классификации – «реальный – виртуальный». Эта шкала располагает агентов по их отношению к объективной реальности, в которой обитаем мы сами. Есть агенты, которые взаимодействуют с ней, а есть и такие, которые существуют в специально построенных для них виртуальных мирах, доступ из которых в наш объективный мир может как существовать, так и отсутствовать. В последнем случае такое виртуальное пространство часто называется «песочницей». Агенты, работающие в нашем мире, называются «роботами». С другой стороны, чисто программные сущности, оперирующие в виртуальных мирах, часто по аналогии называются «софтботами» или просто «ботами».

Вот так можно использовать шкалы классификации. Если есть две дискретные шкалы, то при помощи них можно составить двумерное пространство возможностей.

| **Типы агентов** | | **Реальность** | |
|---|---|---|---|
| | | Реальный мир | Виртуальная реальность |
| **Естественность** | Естественный | Люди | Аватары |
| | Искусственный | Роботы | Софтботы |

Следующая шкала классификации агентов – «неподвижный – мобильный». Например, элементы «умной» транспортной инфраструктуры – это неподвижные агенты, а автономные автомобили – это мобильные агенты, но при этом и тот, и другой типы агентов работают и взаимодействуют друг с другом в рамках одной многоагентной системы. Надо отметить, что эта шкала также ортогональна двум предыдущим, и вместе три шкалы дают возможность описания восьми классов агентов.

Можно выделить ещё большое количество двоичных шкал, по которым можно классифицировать агентов, но большим интересом обладает один из способов классификации на двух ортогональных шкалах. Первая из них определяет, какие средства агент имеет для представления окружающей его среды. Вторая же определяет способ поведения агента.

Первая шкала разносит на два противоположных класса когнитивных и реактивных агентов. Когнитивные агенты обладают более богатым представлением среды по сравнению с реактивными, поскольку у них имеется как механизм логического вывода, так и база знаний с символьным описанием внешней среды. С другой стороны, реактивные агенты либо вообще не имеют внутреннего описания среды, либо оно настолько бедно, что не позволяет им прогнозировать развитие ситуации, а потому они действуют исключительно в рамках процесса «стимул – реакция» (отсюда и наименование). Однако когнитивные агенты могут прогнозировать развитие среды и, стало быть, планировать свои действия в ней.

Вторая шкала отчасти связана с первой и довольно сильно коррелирует с ней, но в то же время допускает отклонения. Она определяет способ поведения агента в окружающей среде, и здесь выделяется континуум значений, которые могут быть разделены на три обширных класса – реактивное, активное и проактивное поведения. Реактивное поведение обусловлено исключительно реакцией агента на стимулы из среды. Активное поведение связано не только с внешними, но и с внутренними стимулами. Проактивное поведение задействует прогнозы и планы, разрабатываемые агентом, и действует упреждающе в отношении изменения среды, равно как и в дальнейшем использует разницу между прогнозом и фактом для обучения.

| **Типы агентов** | | **Тип поведения** | | |
|---|---|---|---|---|
| | | Реактивное | Активное | Проактивное |
| **Представление среды** | Бедное | Растения | Низшие животные | – |
| | Богатое | – | Высшие животные | Люди |

Теперь перейдём к способам классификации самих многоагентных систем. Здесь также можно рассмотреть несколько типов классификаций, однако теперь в качестве основного критерия для классификации берутся различные аспекты многоагентных систем, друг с другом не сравнимые.

Традиционно многоагентные системы классифицируют на три класса.

1. Распределённый искусственный интеллект.
2. Децентрализованный искусственный интеллект.
3. Искусственная жизнь.

В первом случае многоагентная система распределённого искусственного интеллекта чаще всего создаётся динамически для решения какой-то конкретной задачи. Фактически этот подход является развитием имитационного моделирования при помощи агентов, когда создаётся среда, в неё запускаются агенты для решения конкретной задачи. Процесс централизованно координируется главным агентом, который может направлять управленческие воздействия каждому агенту в многоагентной системе. Задача решается распределённо, но результаты решения собираются центральным агентом.

Децентрализованный искусственный интеллект предназначен для того, чтобы исследовать поведение агентов в динамической многоагентной системе во взаимодействии с другими агентами. В этом случае центральный управляющий агент может отсутствовать (чаще всего так и есть), а сбор информации о поведении осуществляется в рамках надсистемы. Децентрализованный искусственный интеллект возникает как результат взаимодействия и взаимного усиления агентами друг друга.

Наконец, парадигма искусственной жизни была введена для проведения исследований и моделирования различных процессов, которые часто встречаются в сложном окружении обычной физической реальности – децентрализованного управления, эволюции, адаптации, конкуренции и кооперации, а также спонтанного возникновения сложности. В составе искусственной жиз-

ни могут быть агенты различной природы и типов, и вопросы их взаимодействия находятся в фокусе внимания исследователей.

Важным понятием теории многоагентных систем является *роевой интеллект*, который эмерджентно возникает в многоагентной системе тогда, когда каждый отдельный агент является простым, часто реактивным и без признаков интеллекта, но сложное поведение и взаимодействие агентов друг с другом и средой позволяет возникнуть новому феномену – интеллекту всей системы. Это наиболее острая тема в рамках изучения многоагентных систем.

Роевой интеллект – это методология, которая описывает коллективное поведение децентрализованных самоорганизующихся систем естественной или искусственной природы. Фактически роевой интеллект как метод моделирования и познания является апофеозом теории многоагентных систем. Обычно система роевого интеллекта состоит из набора агентов (иногда их ещё называют «боидами»), которые взаимодействуют друг с другом и своим окружением. Каждый агент следует простым правилам поведения, но из-за отсутствия централизованного управления интеллектуальное поведение всей системы возникает как эмерджентное свойство, неизвестное и недоступное каждому индивидуальному агенту по отдельности.

Можно рассмотреть такой пример из области транспортных систем. Возьмём общественный пассажирский транспорт. Где здесь могут быть применены технологии многоагентных систем? Ну, к примеру, в задаче обеспечения приоритетного проезда перекрёстков и светофорных объектов подвижными единицами пассажирского транспорта.

Давайте посмотрим на светофорный объект на перекрёстке. Он управляется специальным дорожным контроллером, который может быть рассмотрен в качестве агента одного из типов. Этот агент умеет выполнять важные функции – переключать огни светофоров, менять длительности фаз, включать стрелку на правый поворот с полосы для общественного транспорта и т. д.

Другой тип агента – это подвижная единица пассажирского транспорта, которая по своему маршруту везёт пассажиров. Этот агент всегда знает, сколько людей находится внутри подвижной единицы, а также насколько он отстаёт или опережает график своего маршрута. Кроме того, у этого агента есть ресурсы – то количество денег, которое заплатили воспользовавшиеся его услугами люди.

И вот с конкурирующих направлений к перекрёстку подъезжает несколько автобусов. В одном мало людей, и он не опаздывает

по своему графику. В другом же людей много, и он чуть-чуть опаздывает. Второй агент при приближении к перекрёстку запрашивает у контроллера текущий график переключения фаз, после чего понимает, что если «протянуть» текущую фазу на 30 секунд, то он успеет проехать перекрёсток и немного сократит отставание по графику.

В этот момент агент отправляет на контроллер запрос о продлении фазы на 30 секунд. Контроллер рассылает всем приближающимся подвижным единицам предложение по изменению графика. Второй агент понимает, что если это предложение будет принято и исполнено, то он отстанет от графика на несколько секунд. И тогда он отправляет на контроллер просьбу не принимать новый график, а также сумму, которую он готов заплатить за это.

Надо отметить, что первый агент также послал сумму, которую он готов потратить на то, чтобы контроллер протянул фазу. Контроллер принимает решение с учётом того, что отступление от установленной программы тоже стоит каких-то ресурсов, поэтому в расчёт идёт не только размер предлагаемой всеми приближающимися агентами платы.

Здесь может быть несколько итераций обмена сообщениями между агентами, и тогда может быть устроено что-то типа аукциона, в том числе и с учётом новых агентов, которые могут приблизиться к перекрёстку, пока первоначальные занимались переговорами. Но контроллер может и самостоятельно и сразу принять решение и сообщить об этом всем окружающим агентам, разослав новый график и тот интервал времени, в течение которого изменение графика больше невозможно.

А что происходит вокруг этой системы? А снаружи водитель автобуса подъезжает к перекрёстку и видит, что светофор, который, казалось бы, уже должен переключиться, продолжает гореть для него. Водитель проезжает и следует дальше по своему маршруту. А пассажиры внутри салона вообще ничего не замечают. Вот примерно так и должны работать современные интеллектуальные системы, делая жизнь человека удобнее и лучше.

Сегодня многоагентные системы являются одним из наиболее интересных направлений исследований в рамках искусственного интеллекта и смежных наук. Они применяются для решения задач моделирования, оптимизации и управления в компьютерных играх, на транспорте различной модальности и в логистических системах, в робототехнике разных направлений. Так что всё это очень интересно.

# Глава 4

# ФИЛОСОФИЯ

Рассмотренные технологии построения искусственных интеллектуальных систем позволяют создавать поистине прекрасные творения человеческого разума. Уже сегодня мир людей наполнен автономными ИИ-системами, про функционирование которых большинство людей даже не догадывается. Однако решает ли искусственный интеллект как дисциплина поставленную перед ним изначально цель – моделировать когнитивные функции человеческого разума для изучения этого самого разума, – пока не совсем ясно.

Вместе с тем лично я глубоко убеждён, что любой инженер, программист, математик и любой человек такого же склада ума, когда хочет начать работать над искусственными интеллектуальными системами, должен узнать, как устроен естественный интеллект. Более того, такой человек должен постоянно изучать новую информацию из этой области, потому что мы до конца не знаем об устройстве нашего мозга и нашего сознания, а новые знания здесь появляются с завидной периодичностью.

Но для чего же это нужно? Во-первых, это гуманитарное знание позволит не забывать о том, что любое развитие научно-технического прогресса должно в конечном итоге вести к улучшению жизни людей и прогрессу человеческой цивилизации в целом.

Во-вторых, изучение этих тонких материй – естественного интеллекта и его материального носителя – позволит продвигаться и по пути построения интеллекта искусственного, поскольку пока нам всё равно приходится брать знания у природы. Других интеллектов у нас нет, а потому, если мы хотим создать искусственный, нам надо изучать естественный.

Так что в следующей главе мы погрузимся в философию сознания и философию искусственного интеллекта. На основы нейрофизиологии я замахиваться не стану, так как этой важной дисциплине посвящено достаточно материала.

# Раздел 4.1. Философия сознания

Начиная изучать философию сознания, давайте начнём с главного – определения предмета этой дисциплины. Ведь сказать, что предметом философии сознания является сознание, слишком просто, а фактически это значит ничего не сказать. Более того, это слишком общо, и определить термин «сознание» крайне сложно, ведь иначе, если бы мы его определили, говорить было бы вообще не о чем. Так что давайте решим, что мы будем изучать.

В чём же дело? Дело в том, что у нас есть две субстанции. С одной стороны, у нас есть нервная система человека и её центральный орган – головной мозг. Это самый сложный объект из известных нам во Вселенной. Мозг человека структурно состоит из десятков различных компонентов, каждый из которых представляет собой переплетённые нейронные сети, отправляющие свои сигналы всем другим элементам мозга. И неокортекс, новая кора, – это большое поле из так называемых кортикальных колонок, каждая из которых состоит из нескольких сотен нейронов, а всего таких колонок в новой коре около трёхсот миллионов. Нейроны в неокортексе переплетены друг с другом сложнейшим образом, и есть нейроны, которые обмениваются сигналами с десятками и сотнями тысяч других нейронов. Всего в головном мозге человека имеется порядка девяноста миллиардов нейронов. Всё это великолепие, потребляя примерно двадцать процентов всей энергии организма, работает постоянно, непрерывно и параллельно, обрабатывая гигабайты входящей сенсорной информации в секунду.

Каждый естественный нейрон – это сложная биохимическая машина для обработки информации. Он как будто бы складывает входящие в него импульсы с учётом возбуждающего и тормозного влияний каждого импульса. И если частота пришедших импульсов выше какого-то порога, он сам испускает исходящий импульс, используя для передачи на следующий нейрон специальное химическое вещество (нейромедиатор), которое не только передаёт возбуждение на следующий нейрон, но может также воздействовать и на соседние нейроны, возбуждая или подавляя их. А ещё оно может служить в качестве гормона, осуществляя сигнальную функцию «для всех вокруг». И всё это работает на стохастических принципах молекулярной вычислительной па-

радигмы. Всё это великолепие отвечает за когнитивные процессы в нашем головном мозге.

А с другой стороны, у нас есть сознание. Это нечто, воспринимаемое нами непосредственно в качестве своего «Я», в качестве самого себя. Это очень субъективная штука, и ей сложно даже дать полное определение. Мы можем давать только лишь этакие операционные определения сознанию для решения определённых задач, скорее философского характера. Но дать полного, точного, исчерпывающего определения сознанию невозможно, поскольку его природа пока ускользает от понимания человека.

Ну и, соответственно, главным вопросом, который будет изучаться со всех сторон в этой главе, является вопрос о соотношении биологического и психического. Как в биологической субстанции, которая представляет собой головной мозг во всём его великолепии и сложности, зарождаются психические феномены. Особенно нас будут интересовать среди всего их разнообразия именно сознание, самосознание, восприятие и *квалиа*. Всё это мы внимательнейшим образом изучим далее. Это всё есть проблема соотношения мозга и сознания. Эта проблема и представляет собой сегодня главный предмет изучения философии сознания, и в его изучении уже достигнуты определённые результаты.

Но вот что интересно. В процессе изучения этого вопроса мы натыкаемся на определённого рода методологическую проблему. Со стороны биологической субстанции у нас много чего изучено – мы можем погрузиться до субклеточных уровней и изучать работу отдельных нейронов чуть ли не с позиции наблюдения за отдельными биохимическими реакциями. Объём наших знаний о функционировании нейронов и других клеток нервной системы растёт, как снежный ком. Но и со стороны психологического изучения деятельности человека также есть много достижений. Различные отрасли психоневрологии изучают всё многообразие процессов внутри нервной системы от достаточно простых и низкоуровневых до высших процессов нервной деятельности.

А вот между двумя этими областями и их достижениями существует существенный разрыв в понимании. Есть понимание биохимических механизмов, и есть более или менее сносное понимание механизмов психических, но связи между ними никакой не прослеживается. Собственно, из-за этого разрыва, из-за этой методологической особенности изучения мозга и нервной системы и встаёт проблема мозга и сознания. И пока доволь-

но сложно представить себе, как этот разрыв будет преодолён. Само собой разумеется, что здесь нам потребуется комплексный междисциплинарный подход, в рамках которого будут работать специалисты из самых разных направлений науки.

Перед тем как погружаться в глубины философии, необходимо дать хотя бы операционные определения сознания. Таковых, как минимум, два, поскольку есть два подхода к пониманию сознания даже на бытовом уровне. И для начала давайте посмотрим на узкое определение сознания. Для этого вспомним такие выражения, как «находиться без сознания» или «потерять сознание». Что это обозначает? В принципе, это такой неформальный медицинский термин, который мы понимаем в целом интуитивно. Мы говорим, что человек находится без сознания, если у него не наблюдаются свидетельства высшей нервной деятельности: он не разговаривает, никак не реагирует на большинство внешних стимулов. Кроме того, у него могут отсутствовать важные рефлексы, например зрачковый рефлекс на действие света, т. е. зрачок не сжимается, если светить человеку лампочкой в глаз.

В таком понимании сознания есть несколько проблемных вопросов. Во-первых, сразу хочется отбросить дуальность – сознание есть, и сознания нет. Вполне могут быть промежуточные состояния. Более того, могут быть состояния сознания, которые вообще не находятся на этой шкале, – так называемые изменённые состояния сознания, которые вызваны, к примеру, приёмом некоторых специальных препаратов или какими-либо физиологическими воздействиями. Человек может находиться в вегетативном состоянии, в коме, под общим наркозом, в сопоре, в затуманенном усталостью или голодом сознании, в обычном своём состоянии, в состоянии повышенной, экстатической возбудимости и многих других – это всё разные степени проявления сознания.

Вторая проблема в том, что фраза «некто находится без сознания» может быть применена не только к человеку. Например, мы легко можем применить её к шимпанзе – нашему ближайшему родственнику. Критерии определения состояния будут те же самые. В принципе, такие же критерии легко подходят и к собакам, кошкам, коровам, да и вообще любым млекопитающим. А теперь вспомните жука, который падает в траву с былинки, поджав лапки, когда до былинки кто-то дотронулся. Исследования энтомологов показывают, что в эти моменты жук находится «без сознания». Ну уж если к жуку можно применить это выра-

жение, то почему тогда мы не можем сказать, что отключенный от питания робот находится «без сознания»? Это ведь всего лишь вопрос терминологии. Но если мы согласимся считать нефункционирующего робота находящимся без сознания, то почему мы будем отказывать в этом всем иным техническим устройствам? Да, это действительно проблема.

Ну, хорошо. А каково тогда общее понимание сознания? С этим сложнее. Уже было написано, что как такового определения нет. Ведь если бы оно было, писать эту главу не было бы нужды, да и само направление искусственного интеллекта не зародилось бы. Но в качестве рабочего понимания можно сказать, что сознание в общем смысле – это ощущение «самости», ощущение своей отдельности от окружающего мира, выделение своего «я» и его самосознание.

И вот с этим возникает ещё большая проблема, но уже другого плана. Мы опять же интуитивно понимаем, кто является сознательным, а кто нет. Проблема в том, что чётко показать это мы не можем. Кроме как к самим себе применить своё интуитивное понимание сознания и сознательной деятельности невозможно. Для понимания этого можно поставить такой мысленный эксперимент, который называется «философский зомби». Представим существо, которое выглядит как человек. Оно разговаривает, реагирует на внешние стимулы и всё остальное делает так, как это во всех ситуациях делает обычный человек. Единственное отличие – у этого существа нет той самой субъективной внутренней реальности, оно как будто бы ничего не воспринимает. Оно – зомби, у которого всё темно и тихо внутри. Но при взаимодействии с этим существом вы не сможете ни доказать, ни опровергнуть наличие у него сознания. Если его спросить об этом, это существо скажет, что у него есть субъективный опыт. Ведь оно ведёт себя абсолютно так же, как и обычный человек. Сознание – чисто субъективная штука, поэтому её так сложно изучать научными методами.

Но пока отставим в сторону этот самый «субъективный внутренний мир» как не поддающийся объективному изучению. Давайте подумаем, как можно проверить внешние проявления этого феномена. Если речь идёт об осознании самого себя и отделения самого себя от окружающей среды, то можно подумать в этом направлении. Мы знаем, кто мы такие. Мы можем узнать себя в зеркале. Мы можем сказать, что принадлежит нашему телу, а что нет. Мы можем рассказать о себе истории из биографии. Само

собой разумеется, что сейчас речь ведётся о так называемом «нормальном состоянии» человека, поскольку к любому такому тезису можно привести контрпримеры пациентов с определёнными неврологическими нарушениями, которые не могут делать эти вещи. Но такие нарушения сейчас рассматриваться не будут, хотя в целях дифференциации различных функций сознания и высшей нервной деятельности они очень интересны для изучения.

Так вот узнавание себя в зеркале опять же не является эксклюзивной прерогативой человека. Многие млекопитающие и даже отдельные представители птиц – врановые и большие попугаи – вполне узнают себя в зеркале. Значит ли это, что у них есть ощущение самости? Более того, представим себе робота, у которого есть видеокамера и который запрограммирован таким образом, чтобы как-то специфически реагировать на своё собственное изображение, которое он распознаёт со своей камеры. Разве тогда нельзя будет сказать, что этот робот не обладает самосознанием? Или программная ИИ-система, которая специфически реагирует на предъявляемые ей её собственные машинные коды? Ведь реализовать это проще простого. Реплицирующиеся компьютерные вирусы, которые обрабатывали свой собственный машинный код, были разработаны ещё в прошлом веке. Так, быть может, у таких компьютерных вирусов есть самосознание?

Итак, можно ли сказать про робота или софтбота, который специфическим образом реагирует на своё собственное отображение в его сенсорных стимулах, что он обладает самосознанием? Про робота вы могли подумать, к примеру, что он также специфически может реагировать на любого другого робота, который является точной копией первого. Но, извините, разве вы сами не отреагировали бы особым образом, если бы увидели своего точного двойника только с переставленными правой и левой сторонами, ибо именно такими мы видим сами себя в зеркале? Уверяю вас, ваша реакция была бы более чем специфическая.

В общем, если запрограммировать искусственную интеллектуальную систему таким образом, чтобы она «узнавала» себя, то нет никаких резонов отказывать ей в отсутствии самосознания. Здесь вполне применим так называемый «утиный тест» – если что-то выглядит как утка, плавает как утка и крякает как утка, то с большой степенью уверенности можно заявить, что это утка и есть.

А если возвратиться к тому, каким образом можно было бы реализовать эту функцию узнавания себя, то, с точки зрения про-

граммирования, это банально просто. Не требуется даже иметь в ИИ-системе весь исходный код этой ИИ-системы, достаточно иметь его хеш, и эта система будет сравнивать именно результаты хеширования. Таким образом, мы избавляемся от «дурной бесконечности», которая рекурсивно требовала бы иметь в самой себе копию самой себя.

Но что с животными? Действительно, это очень насущный вопрос – существует ли сознание у животных? Если говорить о сознании в его узком понимании, то несомненно так говорить можно: «Кошка находится без сознания». Так что давайте рассматривать этот вопрос в общем понимании сознания. Тогда ответ на него в большей мере зависит от того, какое операционное общее определение сознания мы принимаем. Например, если принять определение, в котором для описания границ сознания используется понятие второй сигнальной системы, т. е. сложного взаимосвязанного набора символов – конкретных и абстрактных, – то в этом аспекте сознанием обладает только человек и, возможно, некоторые представители высших приматов: шимпанзе, горилл и орангутанов. Если же принимать более слабые определения, то можно дойти и до того, что в какой-то мере сознанием обладает любая биологическая или техническая система, которая может воспринимать стимулы из внешней среды. Хотя, быть может, что-то в этом понимании и есть.

Можно изучить разные возможности. Будем спускаться от наиболее строгого понимания сознания к наиболее расширительному. Первый вариант – это возможность использования символьной системы, что обосновывается гипотезой Ньюэлла–Саймона. Эта гипотеза требует, чтобы разумный субъект с необходимостью обладал возможностью осуществлять символьные манипуляции. В ней, конечно, речь идёт об интеллекте, однако сейчас мы расширяем её и на сознание. То есть пусть сознание – это наличие символьной системы, тогда понятно, что сознательным существом в полной мере можно назвать только человека. А ИИ-систему, построенную на символьном подходе нисходящей парадигмы?

Наши ближайшие родственники – человекообразные обезьяны – могут овладеть языком жестов, и в истории были яркие примеры, например горилла Коко владела примерно двумя тысячами понятий, в том числе и довольно абстрактными, а шимпанзе Уошо применяла 350 жестов. Однако надо понимать, что мы не можем в общем случае опровергнуть существование сим-

вольной системы, например, у котов. Вполне возможно, что мы просто не можем наладить с ними канал коммуникаций, как, например, с дельфинами или слонами, головной мозг которых очень развит. Хотя, конечно, это слабая гипотеза.

Второй вариант – необходимо отделять себя, свою «самость» от окружающей среды. Проще всего определить это при помощи уже рассматривавшегося зеркального теста. Существо смотрит в зеркало и узнаёт в нём себя. Или не узнаёт, а воспринимает изображение как другое существо. Люди начинают узнавать себя в зеркале в возрасте примерно полутора лет. А вот животные имеют различные способности к этому. Доподлинно известно, что этот тест проходят все большие человекообразные обезьяны, дельфины, слоны, а также врановые птицы и большие попугаи. Интересно то, что есть свидетельства прохождения зеркального теста муравьями. Это единственные беспозвоночные, кто смог это сделать, но эксперимент требует проверки.

Третий вариант самый обширный. Это просто возможность сенсорного восприятия стимулов, приходящих из внешней среды. Дело в том, что такое восприятие у человека формирует тот самый опыт восприятия, квалиа, который и характеризует наличие сознания. Мы внутренне знаем, как это иметь такой опыт – воспринимать. Однако, конечно же, это чисто субъективный опыт, и мы не можем знать, есть ли такой опыт у других существ, в том числе и у других людей. В этом-то как раз и заключается тайна природы сознания, её трудная проблема, о которой рассказано далее. Но в данном аспекте получается, что воспринимающим сознанием обладает любое живое существо.

Ну а теперь давайте рассмотрим все три возможности в применении к искусственным интеллектуальным системам, которые находятся в центре нашего интереса. Первая возможность, само собой, присутствует в символьных ИИ-системах, выполненных в рамках нисходящей или гибридной парадигмы. Про вторую возможность уже было сказано чуть ранее – нет ничего сложного, чтобы алгоритмически научить искусственную техническую систему «узнавать» себя. Ну, а про третью возможность и говорить незачем – любая кибернетическая машина так или иначе воспринимает стимулы из среды. Так что даже чат-бот, принимающий сообщения в мессенджере, воспринимает их в качестве внешних стимулов.

Тем не менее вопрос, у кого может быть внутренний субъективный опыт сознательного восприятия, остаётся открытым.

С очевидностью можно говорить о людях – этот опыт непосредственно дан каждому из нас, и отрицать его, называя «иллюзией», контрпродуктивно.

Итак, вопрос о сознании и его природе будоражил умы философов ещё со времён Древней Греции. Если исходить из довольно простого операционного принципа, который предписывает нам объявлять сознательным существо, с которым мы можем так или иначе взаимодействовать на символьном уровне, то с этой точки зрения довольно просто определять, находится человек в сознании или нет. Помните про узкое понимание сознания? Вот сейчас в наших рассуждениях его вполне достаточно. Так вот уже древнегреческие философы, стоявшие на позициях материализма, вполне могли локализовать сознание на уровне органа. Всем было понятно, что сознание находится в голове человека, а если ещё точнее, то в головном мозге.

Далее уже с развитием анатомии и физиологии становилось понятно, что сама нервная система и головной мозг как орган состоят из огромного количества подсистем и элементов. Оказалось, что головной мозг вообще является самым сложным объектом во Вселенной среди известных людям. И это не шутка – именно головной мозг состоит из сотен элементов, связи между которыми переплетены так, что до сих пор не удалось построить более или менее целостную картину их соединения. В каждом элементе могут быть десятки, сотни и тысячи ядер, т. е. скоплений нейронов. А всего в головном мозге примерно девяносто миллиардов нейронов, каждый из которых может соединяться с десятками тысяч других. А некоторые имеют связи с сотней тысяч других нейронов. Если взять всё множество вычислительной техники на Земле, имеющейся у человечества, то при помощи неё невозможно эмулировать связи головного мозга всего лишь одного человека. И при этом мозг взрослого человека весит примерно полтора килограмма и потребляет примерно двадцать пять ватт энергии.

Изучение соотношения мозга и сознания с кибернетической точки зрения с использованием методологии чёрного ящика довольно непросто, так как наталкивается на многочисленные этические проблемы. Невозможно просто так взять и ставить эксперименты на головном мозге живого человека. Приходится искать особые патологические случаи, в которых тщательно разбираться и выявлять те или иные закономерности, к примеру

если у какого-либо пациента неврологической клиники имеются определённые нарушения в сознании, то при помощи электроэнцефалограммы или функциональной магнитно-резонансной томографии можно найти, в каких частях мозга у него имеются проблемы. И наоборот, если есть какие-то проблемы в результатах измерений, то их пытаются сопоставить с когнитивными и функциональными нарушениями в сфере сознания и вообще высшей нервной деятельности. Однако всё это очень нечётко, и, несмотря на то что собран большой объём данных, точных соотношений нет, и сегодня можно говорить только о наличии некоторых корреляций между функциями мозга и его частями.

Например, в нескольких экспериментах на базе многочисленных исследований деятельности мозга была обнаружена высокая корреляция между сознательным состоянием и нейрональной активностью в так называемой ограде мозга – небольшой пластинке нейронов, входящей в состав лимбической системы и тесно примыкающей к неокортексу. Замечательная особенность этой части головного мозга состоит в том, что её нейроны подключены практически ко всем остальным частям. Кроме того, при электрической стимуляции ограды у людей выключается сознание. Однако в конечном итоге были обнаружены пациенты с повреждённой и неработающей оградой, которые внешне создавали впечатление вполне разумных и сознательных людей, а их описание своего субъективного опыта не давало повода предположить, что у них с ним какие-то проблемы. В общем, опять всё сложно.

В итоге в настоящее время главенствующими теориями о том, где находится в головном мозге сознание, являются такие, которые основаны на двух важных научных принципах – супервентности и каузальной замкнутости физического. Супервентность означает, что состояние одной системы детерминируется состоянием другой системы. Например, если у нас есть две вычислительные системы, и у них одинаковые свойства, то эта одинаковость определяет одинаковые ментальные состояния, в которых пребывают эти физические системы. При этом сами ментальные состояния не могут и не должны сводиться к физическим состояниям. Ментальные состояния находятся именно в отношении супервентности к физическим. Каузальная замкнутость обозначает, что любое событие или явление в нашей реальности имеет определённую причину в нашей же реальности. Другими словами, принцип каузальной замкнутости обозначает, что «чудес не бы-

вает». Фактически это основа физикализма, который и является доминирующей основой в современной философии сознания.

Лично мне нравится семейство так называемых К-теорий сознания (К – от слова «когнитивный»). В их рамках рассматривается четыре уровня организации нейронов в коре головного мозга. На самом нижнем уровне находятся отдельные нейроны. Каждый нейрон всего лишь обрабатывает входные сигналы, и эта механика с биохимией очень неплохо изучена. На втором уровне нейроны объединяются в так называемые квалоны, причём один нейрон вполне может состоять в нескольких квалонах. Квалон – это мельчайшая единица восприятия. Если активируется квалон, субъект воспринимает квалиа, т. е. то самое внутреннее субъективное состояние, которое так сложно описать. На третьем уровне квалоны объединяются в опероны. Оперон отвечает не только за восприятие более сложной картины, но и за реакцию на такое восприятие. Наконец, на четвёртом уровне находятся холоны – это объединения оперонов, которые определяют некоторое целостное качественное состояние когнитивной системы.

В соответствии с этими теориями сознание определяется как вот это самое целостное качественное состояние нашего мозга. И сознание как бы распределено по всему неокортексу. Различные состояния сознания и восприятия задействуют и активируют различные области мозга. И то, что мы воспринимаем как так называемый «поток сознания» это последовательная активация холонов. Другими словами, сознание похоже на программное обеспечение, которое работает на аппаратном обеспечении головного мозга. И как в программном обеспечении компьютера есть так называемая точка исполнения, так и у нас в голове есть что-то типа точки внимания. Хотя, конечно, пока это всё гипотезы.

Есть и другие гипотезы, которые пытаются объяснить сознание. Например, есть такие гипотезы, как квантовое сознание и панпсихизм. Гипотезы о квантовом сознании предлагают описывать этот ускользающий феномен через квантовые эффекты. Однако пока не удалось даже спроектировать более или менее стоящий эксперимент для проверки. Панпсихизм выводит сознание за грань физики, объявляя его неким общим свойством материи, имманентно присущим всему. То есть вот скажем, элементарная частица, например электрон, внутренне «знает», что значит «быть электроном». Такое же состояние протона складывается из этих состояний кварков, из которых протон состоит. Ну и сознание че-

ловека равно совокупности таких вот состояний всего того, из чего человек состоит. Красивая гипотеза, но вряд ли верифицируемая. К тому же… Ну почему в каких-то состояниях у человека сознания нет? А когда человек умирает, куда оно девается?

Для понимания дальнейшего изложения давайте кратко охарактеризуем методы исследования активности головного мозга, которые сегодня применяются к человеку. Все такие методы можно разделить на две большие группы – инвазивные и неинвазивные. Инвазивные методы, как следует из их названия, предполагают непосредственное влияние на внутреннее строение головного мозга, и их использование крайне затруднительно. К примеру, к таким методам относится снятие электрических потенциалов непосредственно с коры головного мозга во время трепанации черепа. Вы сами понимаете, насколько это непростая процедура для проведения исследований, а потому применяется крайне редко при проведении операций с пациентами, согласившимися на дополнительные исследования.

К неинвазивным методам исследования относятся такие, которые никак не влияют на деятельность головного мозга. Сегодня наиболее широко используемыми методами являются электроэнцефалография и функциональная магнитно-резонансная томография. ЭЭГ – это съём электрических потенциалов с поверхности черепа. Очень простой метод, который даже позволяет в некоторых случаях диагностировать определённые заболевания, прогнозировать развитие разного рода состояний и даже при помощи мыслей управлять различными устройствами и системами.

Другой метод – фМРТ – намного сложнее для применения, но при этом и намного точнее. Человек сначала принимает специальный препарат, а затем при помощи аппарата МРТ визуализируется динамика изменения потребления глюкозы и кислорода клетками головного мозга. Если какие-либо нейроны, группы нейронов, области коры и подкорковые структуры активно работают при выполнении каких-либо физических или ментальных действий, это хорошо видно в динамике на трёхмерной картинке фМРТ. Работающие нейроны потребляют огромное количество энергии, и именно к ним кровь подводит кислород и питательные вещества, что хорошо видно на диаграмме фМРТ.

При постановке эксперимента предполагается, что в результатах ЭЭГ или фМРТ можно найти нечто, что соответствует сознательному или какому-либо иному функциональному состоянию

человека. Например, человек лежит в аппарате МРТ и пытается планировать своё следующее действие. Картинка деятельности его мозга снимается. Потом тот же человек пытается не планировать своё следующее действие. Опять картинка мозга снимается. Потом из первой картинки как бы вычитается вторая, и полученное изображение называется «коррелятом» того действия, которое делал человек, т. е. планирования. Несмотря на то что это всё очень смутно и нечётко, при постановке большого числа экспериментов с разными людьми можно более или менее определить, какие части коры головного мозга активны, а какие не так активны при тех или иных действиях или ментальных состояниях.

Оказалось, что сознательное состояние связано с высочайшей активностью лобных долей мозга. Нет, другие области тоже активны в те моменты, когда мы переживаем сознательный опыт, но именно лобные доли активны мощнее всего. Это связано, как предполагается, с тем, что в лобных долях осуществляется моделирование, прогнозирование и планирование будущего, и именно поэтому у человека эта область мозга развита лучше всего. С функциональной точки зрения «находиться в сознании» – это значит постоянно моделировать будущее и сравнивать прогноз с фактом, адаптивно изменяя своё поведение. Эта простая формулировка даже наталкивает некоторых исследователей на то, что сознание – это и есть прогнозирование. Впрочем, это очень редукционистская точка зрения.

Однако в сознательных состояниях мозг человека активен не только в лобных долях. Активна вся новая кора и большинство подкорковых структур, причём складывается такое ощущение, что эта активность какая-то беспорядочная, хаотичная. Активность нейронов как будто бы перемещается по всей поверхности коры, иногда угасая, иногда зарождаясь вновь. И в некоторых теориях считается, что вот эта динамическая картина активности нейронов неокортекса как раз и является тем, что мы знаем как «сознание». То есть наше сознание в аппаратном обеспечении головного мозга отображается вот таким динамическим способом.

Эти результаты исследований и размышления наталкивают на мысль, что сознание – это программное обеспечение мозга. Как в компьютерах программное обеспечение является этакой неосязаемой сущностью, так и сознание в нервной системе также представляет собой функции, обусловленные действием биохимических процессов. Однако такое понимание, хотя и является впол-

не приемлемым с точки зрения описания когнитивных процессов, наталкивается на некоторые сложно преодолимые трудности.

Во-первых, с точки зрения теории информации программное обеспечение – это тоже информация, которая должна где-то храниться. В компьютере программа запускается, и из долговременной памяти она загружается в оперативную, где начинается её исполнение. В нервной системе человека нет никаких аналогичных процессов, и совершенно не ясно, где информация обо всём «программном обеспечении» сознания могла бы храниться в головном мозге.

Во-вторых, это понимание сознания как программного обеспечения всё равно не объясняет самую проблемную часть философии сознания – как нечто, чем бы оно ни было, хоть бы и программным обеспечением мозга, может иметь субъективный опыт. Это и есть та самая трудная проблема сознания, про которую многие говорят. Ну а ввёл трудную проблему сознания в научный обиход австралийский философ Дэвид Чалмерс в своей работе «Навстречу проблеме сознания» в 1995 году. Эта проблема так трудна, что её даже трудно объяснить. Но попробуем…

Давайте начнём с того, что в философии сознания выделяется два типа сознания – когнитивное и феноменальное. Когнитивное сознание, или, как его ещё называют, психологическое сознание, – это как раз и есть набор взаимодействующих высших психологических функций. Фактически это совокупность восприятия, осознания, памяти, мышления, умения обучаться, осведомлённости и даже самоосознания. Как ни странно, объяснение всех этих функций не представляет трудности. Мы можем объяснить природу каждого из явлений, которые были перечислены, как в психологических терминах, так и в терминах биохимии.

Например, что такое восприятие? Для примера рассмотрим восприятие различных цветов. Как известно из физики, цвет – это длина волны фотона. Фотон отражается от какой-то поверхности, получает определённую длину волны, после чего попадает к нам в глаз. Далее он реагирует с определённым белком внутри фоторецептора, причём длина волны фотона определяет, с каким именно белком может прореагировать фотон. В результате такого взаимодействия трёхмерная структура белка меняется, запускается каскад биохимических реакций, результатом которых, в свою очередь, является старт потенциала действия от рецептора к первому нейрону и далее в нейроны зрительной коры головного мозга. После этот сигнал распространяется по нейронной сети,

и вся совокупность сигналов от фоторецепторов приводит к восприятию картинки перед нашими глазами. Узнавание происходит на высших уровнях зрительной коры, откуда сигналы далее идут в ассоциативную кору, где осуществляется первичное решение о том, что делать дальше по поводу увиденного.

Собственно, всё предыдущее изложение и было направлено на изучение различных аспектов когнитивного сознания, лишь изредка дотрагиваясь до феноменального, не предполагая этого. Но что же это такое и зачем философам потребовалось вводить новую сущность? Под феноменальным сознанием понимается тот субъективный опыт, который каждый из нас ощущает, испытывает непосредственно. Мы, каждый из нас, в процессе восприятия, мышления, мечтаний и даже при ощущении галлюцинаций испытываем нечто субъективное, вот этот вот опыт осознания. К сожалению, у нас в языке практически нет средств для выражения и описания этого опыта, поэтому вы видите, с какими сложностями мне приходится сталкиваться, объясняя данное понятие. Но вы прямо сейчас испытываете этот опыт, а потому находитесь в состоянии феноменального сознания.

Томас Нагель, ещё один философ, работающий в области исследования сознания, представил свой знаменитый теперь мысленный эксперимент под названием «Каково это – быть летучей мышью». Суть этого важного вопроса в том, что мы как учёные можем досконально изучить всё, что касается биохимии, физиологии и общей жизнедеятельности летучих мышей. Мы воистину можем окончательно познать, «что такое летучая мышь». Но мы никогда не сможем ответить на вопрос «Что значит быть летучей мышью». Потому что на первый вопрос мы можем ответить, исследуя объект от третьего лица. А ответить на второй вопрос возможно, только будучи субъектом и говоря от первого лица. Не являясь летучей мышью, невозможно ответить на этот вопрос. Так вот довольно распространённое и принимаемое многими философами и учёными понимание феноменального сознания заключается как раз в том, чтобы знать, каково это – быть сознательным существом. Феноменальное сознание – это обладание субъективным опытом, что такое сознательное восприятие окружающей реальности.

Давайте рассмотрим различные модальности сознательного опыта, который доступен нам, людям. Первое, что приходит на ум, – это визуальный опыт. Цвета, формы, размеры, яркость, глубина – это всё различные свойства визуального феноменального

опыта, который доступен нам. Слуховой опыт тоже имеет разные свойства: громкость, тональность, музыкальность. Тактильный опыт разделяется на такие свойства, как шершавость, гладкость и вообще разная фактура, различная температура, боль и другие ощущения от прикосновений. Обонятельный опыт вообще самый невыразимый – у нас в языке просто нет слов для названия разных запахов. Также и вкусовой опыт довольно сложно описать, хотя несколько слов для этого у нас есть. Есть много внутренних ощущений, которые также лежат в основе субъективного опыта, – вестибулярные ощущения, голод и насыщение, дрожь, зуд, щекотка, позывы к мочеиспусканию и даже оргазм. Вот попробуйте описать словами оргазм как внутреннее ощущение, и вы поймёте, насколько это непросто. Но субъективный опыт есть. Кроме всего перечисленного, у нас есть такие субъективные ощущения, как ментальные образы, эмоции, чувства и осознанное мышление. Всё это у нас где-то внутри, и немногое из этого мы можем описать. Трудная проблема сознания заключается в ответе на вопрос, почему в том или ином физическом объекте внезапно появляется вот такой вот субъективный опыт.

Для описания этого субъективного опыта используется понятие квалиа. Квалиа – это осознаваемое качество сенсорных сигналов. Давайте подробно разберёмся в этом определении, рассматривая каждое слово с самого конца.

Итак, осознаваемое качество сенсорных сигналов. Сигналы – это то, что наш организм воспринимает из внешней или внутренней среды. Это огромное количество сигналов различной природы – фотоны, акустические волны, сила тяжести, воздействие теплового движения молекул, сами молекулы в качестве сигналов. В каждый момент времени наш организм просто бомбардируется мощнейшим потоком этих сигналов. И в этом нет ничего необычного – каждый такой сигнал вполне может быть описан в физических терминах.

Идём дальше. Почему нас интересуют только сенсорные сигналы? Дело в том, что, к примеру, фотон, попавший к нам на кожу, может рассматриваться как сигнал, пришедший в наш организм, но эффект имеет только фотон, который попал на сетчатку глаза и прореагировал там со специальным белком в фоторецепторе. Также и с остальными типами сигналов – они должны попасть в определенный сенсор нашего организма. Сенсор всегда переводит пришедший из среды сигнал в нервный импульс. Этот

процесс также может быть достаточно просто описан в биохимических терминах.

Продолжаем. Что такое качество сенсорного сигнала? Вот с этим уже сложнее. Под качеством мы понимаем, скорее, модальность – т. е. то самое ощущение, которое возбуждает сенсорный сигнал. Именно это качество мы описываем на нашем обыденном языке такими терминами, как зрение, слух, вкус, обоняние, осязание и т. д. Наша нервная система получает импульсы с различных сенсоров и как бы классифицирует их по качествам. Зрительные сигналы обрабатываются в одних зонах коры головного мозга, обонятельные сигналы – в других. В общем, можно считать, что качество сенсорного сигнала так или иначе соотносится с некоторыми определёнными зонами неокортекса и другими элементами нервной системы. Хотя, честно говоря, вот это уже довольно слабо. Таким образом, качества сенсорных сигналов мы более или менее можем описать при помощи терминов нейрофизиологии.

И наконец, что же такое осознаваемое качество сенсорных сигналов? Вот с этим-то как раз и возникают самые сложные вопросы. Что значит «осознаваемый»? Это значит, что вот я, кем бы я ни был, субъективно воспринимаю и осознаю некоторое качество сенсорного сигнала. Объяснить это действительно непросто, но каждый это делает ежемоментно. Так что рассмотрим это на некоторых примерах.

Мы знаем, что у нас в сетчатке глаза имеются фоторецепторы четырёх типов. У некоторых людей – пяти типов, но это счастливчики, и мы к ним ещё вернёмся. Так вот, один тип фоторецепторов – это палочки, которые воспринимают интенсивность света, и сейчас они нам не очень интересны. Другие три типа рецепторов воспринимают красный, синий и зелёный цвета, то есть, как ранее разъяснено, специальные белки в них реагируют с фотонами разных длин волн. Другими словами, у нас сетчатка устроена по принципу RGB, в ней рецепторы воспринимают только три базовых цвета. А теперь возьмём, скажем, фиолетовый цвет. Для любого из нас, кроме страдающих дальтонизмом, очевидно, что мы «видим» фиолетовое. Но в нашей нервной системе нет ничего фиолетового, там есть только нейроны, которые передают друг другу импульсы. Вот это появляющееся ощущение фиолетового у нас внутри – вот это и есть квалиа.

Более того, квалиа не обязательно должны появляться в ответ на сенсорные стимулы из внешней среды. Представьте ситуацию,

что вы сидите и краем глаза видите фиолетовое пятно на стене слева от себя. Вы поворачиваете голову, но никакого пятна там нет. Но качественное ощущение фиолетового у вас же было? Ведь было же! Хотя ни один из синих и красных фоторецепторов в сетчатке даже не активировался. То есть квалиа – это нечто совершенно субъективное.

Вспомните мысленный эксперимент под названием «Что значит быть летучей мышью». Мы вряд ли сможем узнать, как же всё-таки чувствовать себя летучей мышью. Ну просто потому, что у нас с летучими мышами нет каналов коммуникаций, даже если предположить, что они разумны. Но мы можем провести такие исследования с другими людьми. Есть такие счастливые женщины – тетрахроматы. Их около двух процентов от всего населения, и это только женщины. Так вот у них в сетчатке четыре типа цветовых фоторецепторов. А поскольку это люди, то мы с ними можем пообщаться и постараться узнать, как это – «быть тетрахроматом». Проблема в том, что они сами, скорее всего, даже не знают, как это – быть тетрахроматом, поскольку у человечества нет в языке специальных слов для обозначения этих квалиа, которые воспринимают эти счастливицы. Но с теми из них, кто задумывается о том, почему два разных синих цвета во всех языках мира называются одинаково, можно об этом поговорить и постараться их понять.

Теперь давайте проведём такой мысленный эксперимент, который позволит приблизиться к интуитивному пониманию квалиа. Эксперимент называется «Комната Марии». Представим себе Марию, великолепного учёного, которая всю свою сознательную жизнь посвятила изучению вопроса цветового восприятия у людей. Она изучила всю литературу по этой проблеме. Она знает физику и биохимию процесса. Она знает, что красный цвет – это восприятие фотонов с длиной волны от 625 до 760 нм. Но у неё есть единственная проблема – она заперта в комнате, где всё вокруг чёрно-белое. Она видит вокруг себя только одни градации серого цвета, хотя с её глазами всё в порядке. Представим, что её тело тоже каким-то чудесным образом отвечает этому ограничению, т. е. Мария не знает, что значит субъективно воспринимать красный цвет. Хотя она знает о нём всё. Подумайте над этим. Хорошенько подумайте, а потом ответьте на такой вопрос: «Что будет, если Марию выпустить из комнаты и показать ей красное платье?»

Итак, что увидит Мария, когда ей покажут красное платье? Какое квалиа она испытает? Испытает ли она его? Ответа на этот

вопрос нет, так как квалиа – это субъективный опыт, и мы не можем изучить его от третьего лица, только от первого. А это значит, что опыт другого человека будет для нас неизвестен. Поэтому мы можем только спекулировать на тему того, что увидит Мария. Некоторые говорят, что у неё будет что-то типа «слепоты невнимания», она просто не сможет воспринять красный цвет платья, так как у неё для этого нет в неокортексе нейронных связей. Иные говорят, что она воспримет квалиа красного цвета, но не сможет выразить это и не сможет даже сама себе объяснить, что же она увидела. Однако произойдёт что-то новое для неё.

Теперь попробуем изучить место сознания в окружении так называемых высших психологических функций. Это понятие ввёл Лев Семёнович Выготский, который обозначил этим термином наиболее сложноорганизованные психофизиологические процессы. Согласно Выготскому, высшие психологические функции возникают на основе «натуральных психологических функций» за счёт опосредствования их «психологическими орудиями», например знаками или символами. И соответственно, к высшим психологическим функциям относятся восприятие, память, мышление, воображение и речь.

Давайте кратко рассмотрим каждую из перечисленных функций. Первая функция – восприятие. Это познание объективной реальности при помощи органов чувств. Можно выделить непосредственное чувственное восприятие, и таковое имеется как у нас, людей, так и фактически у всех живых существ. Но также можно выделить опосредованное предметное восприятие, когда мы субъективно опредмечиваем среду, т. е. выделяем в ней различные объекты и явления, давая им различные наименования. Это может делать только человек, насколько нам известно. Через восприятие мы познаём окружающий мир. И, как вы понимаете, без специальных приборов мы не можем воспринимать объективную реальность вне диапазонов восприятия наших органов чувств.

Вторая функция – память. Фактически это познавательная способность по накоплению, сохранению и воспроизведению знаний и навыков. Память в том или ином виде имеется у многих животных и даже растений, но только человек обладает наиболее развитым уровнем памяти, которая к тому же вынесена вовне – на внешние носители памяти. Ранее это были только письменные записи, но сегодня это и огромное количество цифровых носителей информации.

Идём дальше, и третья функция – это мышление. Фактически это высшая ступень процесса познания окружающего реального мира, основу которого составляют образование и непрерывное пополнение запаса понятий и представлений. Мышление включает в себя вывод новых суждений и осуществление умозаключений. Также мышление позволяет получить знание о таких объектах, свойствах и отношениях объективной реальности, которые не могут быть непосредственно восприняты при помощи органов чувств. Предполагается, что мышлением в зачаточном виде обладают только некоторые высшие животные, а полноценно, конечно же, человек.

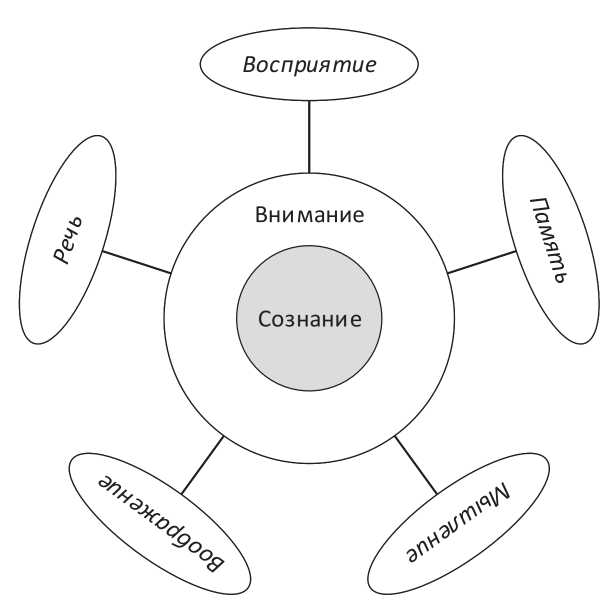
Продолжим, четвертая функция – воображение. Это способность человека к спонтанному возникновению или преднамеренному построению в сознании образов, представлений и идей, которые в опыте в целостном виде не воспринимались или не могут восприниматься посредством органов чувств. Например, это могут быть события истории предполагаемого будущего, явления несуществующего мира – сверхъестественные персонажи сказок или мифов. Воображение лежит в основе моделирования, планирования, творчества и игры. И похоже, что воображение является прерогативой человека.

Наконец, пятая функция – это речь. Под речью мы понимаем существенную функцию человека, позволяющую нам познавать окружающий мир, передавать свои знания и опыт другим людям, аккумулировать их для передачи последующим поколениям. Фактически только речевая деятельность даёт нам, людям, возможность отвлечения от действительности и обобщения, что и является отличительной особенностью человеческого мышления. Речь – это форма общения людей посредством языка, а язык – это вторая сигнальная система, изначально присущая только человеку.

Что интересно, все перечисленные высшие психологические функции могут осуществляться и без сознания. Восприятие и память – очевидные функции, которым сознание не требуется. Измерительный прибор обладает функцией восприятия какого-либо параметра окружающей реальности. У компьютера есть память. С мышлением, воображением и речью сложнее. Если речь можно так или иначе эмулировать на компьютере, то мышление и воображение – это исключительная прерогатива высших животных и человека. Так что в этом отношении у нас ещё долгий путь к полной эмуляции всех высших психологических функций.

То есть сознание само по себе не является причиной этих функций. Но и набор этих функций в одном существе не требует наличия у него сознания. Мы уже рассматривали философского зомби, который ведёт себя абсолютно так же, как человек, но при этом не обладает субъективным опытом. Если сознание – это набор квалиа, то у философского зомби сознания нет. Но внешние его проявления таковы, что мы не может отказать ему в наличии восприятия, памяти, мышления, воображения и речи. То есть получается, что высшие психологические функции не определяют сознание и оно не является их причиной. Они находятся в неком ином соотношении.

Я предполагаю, что сознание связано с каждой из функций через внимание, через фокусировку внимания на функции. Например, мы ежесекундно воспринимаем огромное количество сигналов, но фильтр таламуса в головном мозге пропускает до сознания только то, что необходимо в текущий момент. Или мы можем сознательно переключить наше внимание на восприятие, скажем, звуков. С остальными функциями то же самое. Сознательно мы можем обратиться к памяти, осмыслять определённую идею, мечтать об определённых объектах и использовать речь сознательно.



В общем, сознание находится в центре высших психологических функций, при этом оно соотносится с ними через фокусировку внимания. Это надо хорошо понимать, потому что сознание не является ни достаточным, ни необходимым фактором наличия той или иной функции.

Сейчас давайте сфокусируемся на пятой высшей психологической функции – речи, вернее, на производной этой функции – второй сигнальной системе. Посмотрим, как она формирует наше восприятие объективной реальности.

Понятие второй сигнальной системы ввёл в научный оборот великий русский физиолог Иван Петрович Павлов. Вообще, он говорил о первой и второй сигнальных системах у животных и человека, а само понятие сигнальной системы обозначает набор безусловных и условных рефлексов в нервной системе живого существа. Сигнальная система воспринимает параметр окружающей реальности при помощи сенсоров и генерирует на основе воспринятого сигнал, который обрабатывается внутри системы и приводит к возникновению той или иной реакции. Это и есть тот самый кибернетический подход, который принят для описания подобных феноменов.

Первая сигнальная система присуща практически всем живым существам. Можно даже описать рефлексоподобные реакции у одноклеточных организмов и растений, у которых в принципе нет нервной системы. Но они тоже могут воспринимать различные сигналы – химические, световые, звуковые и даже тактильные. Возможно, что некоторые существа могут воспринимать сигналы совершенно непостижимой для нас модальности, например магнитное поле Земли. Но животные точно обладают первой сигнальной системой, и это всё, что животное воспринимает, весь комплекс квалиа, которые формируются у него в нервной системе. Это то, что животное имеет в себе субъективно, все впечатления, ощущения и представления об окружающей внешней среде. И у нас, людей, тоже есть эта сигнальная система.

Вторая сигнальная система намного сложнее. Это, как писал Павлов, система «сигналов сигналов», т. е. символьное представление окружающей реальности. Речь человека – это важнейшая и главная часть второй сигнальной системы. Именно речевые рефлекторные контуры в головном мозге человека и формируют вторую сигнальную систему. Именно речевые символы «*представляют собой отвлечение от действительности и допускают обобще-*

*ние, что и составляет наше личное, специально человеческое, высшее мышление*». Другими словами, при помощи второй сигнальной системы мы объясняем себе окружающую действительность.

Вторая сигнальная система абстрагирует реальность и даёт нам возможность использовать исключительно внутреннее представление для осмысления и познания. Нигде в объективном мире мы не найдём таких понятий, как «красота», «свобода», «мысль», «категория», «система» и т. д. Огромное количество абстрактных понятий сформировано человеком в рамках второй сигнальной системы исключительно для того, чтобы познавать окружающую действительность. Более того, мы можем использовать её для обмена информацией и организации совместной познавательной деятельности. Именно этот коллективный процесс позволяет человечеству двигаться вперёд по пути научно-технического прогресса.

Также мы используем символы для постоянного «объяснения мира», т. е. описания объектов и явлений, которые нас окружают. Мы делаем это для себя, мы делаем это совместно с другими людьми. Это основа так называемого построения модели реальности, которое бессознательно постоянно происходит у нас в голове. Вот, например, мы видим, как некоторое животное перебегает дорогу. На вход нашей зрительной сенсорной системе приходит последовательность фотонов, возбуждающих фоторецепторы и передаточные нейроны, которые затем формируют некоторые возбуждения в зрительной коре головного мозга. Далее происходит распознавание и активируются рефлекторные контуры между зрительной и ассоциативной корой, что возбуждает комплексы нейронов в коре головного мозга, ответственные за такие символы, как «чёрное», «кошка», «млекопитающее», «животное», и многие другие ассоциации, полный набор которых зависит от конкретного человека. И вот здесь имеет место именно символьное представление информации – мы описываем себе произошедшее событие при помощи слов. И далее могут возбудиться даже вторичные и более высоких порядков рефлекторные контуры, которые основаны на культурном восприятии ситуации. Человек, увидев чёрную кошку, которая перебежала ему дорогу, может повернуть назад. И далее он также при помощи символов, т. е. слов, объяснит самому себе, почему он это сделал.

Почему это так важно для специалистов, изучающих искусственный интеллект? Дело в том, что в области разработки искусственного интеллекта в далёком 1976 году Аллен Ньюэлл и Гер-

берт Саймон сформулировали гипотезу, которая впоследствии получила их имя. Эта гипотеза гласит: «*Физическая символьная система имеет необходимые и достаточные средства для произведения базовых интеллектуальных действий в широком смысле*». Другими словами, без символьных вычислений невозможно выполнять осмысленные действия, а способности выполнять символьные вычисления вполне достаточно для того, чтобы стать способным выполнять осмысленные действия. А сами символьные вычисления – это есть манипуляция символами, т. е. использование второй сигнальной системы для обработки символьной информации. Таким образом, в соответствии с этой гипотезой использование символьной системы при обработке информации является необходимым и достаточным критерием отнесения системы к интеллектуальным. Гипотезу сложно подтвердить или опровергнуть, но, как минимум, один подтверждающий пример у нас есть – это мы сами.

В общем, интеллект связан с символами и их обработкой. То, что у людей на базовом уровне это делают нейронные сети, не имеет никакого значения. Вполне вероятно, что и в ИИ-системах это тоже будут делать искусственные нейронные сети, но может случиться и так, что искусственный разум зародится в гибридных схемах, в которых символьные вычисления будут осуществляться логическими машинами вывода. Но то, что у искусственной интеллектуальной системы в сильном её понимании должна быть вторая сигнальная система, это важная гипотеза, которой необходимо уделить самое пристальное внимание.

Теперь можно подумать вот о чём. После Ивана Петровича Павлова некоторые исследователи ввели понятие «третьей сигнальной системы». Некоторые понимают под ней набор квалиа (что довольно странно, так как плодит сущности без необходимости), иные утверждают, что это система обмена информацией между людьми на уровне идей и образов без необходимости облекать их в словесное выражение. Идеи и образы – это тоже ведь символы, но более высокого порядка. Подумайте над этим и попытайтесь ответить на вопрос, возможна ли третья сигнальная система в этом понимании у людей? А у людей и сильных ИИ-систем?

Конечно, однозначных ответов тут быть не может, так как мы многого не знаем. Но вот я считаю, что прямой связи между людьми для непосредственной передачи идей и образов быть не может. Смотрите. Мы – это наш коннектом, т. е. весь дина-

мический набор связей между нейронами. Даже у генетически тождественных близнецов коннектомы разные, а это намекает на то, что именно коннектом определяет нашу личность и, как следствие, память. В памяти же так или иначе хранятся образы, к которым мы обращаемся при размышлениях. Образ есть последовательность возбуждения определённых нейронных комплексов, соединённых специфичным образом. А эта специфичность у каждого своя, и если мы каким-то образом сможем передать эту последовательность возбуждений от одного человека к другому, то у второго просто не будет тех же самых нейронов и их соединений, чтобы переданную информацию воспринять должным образом. Она может случайно наложиться на существующие нейроны и связи, и тогда второй человек ощутит нечто хаотичное. Но мы, конечно, можем сделать это опосредованно, если между людьми будет находиться устройство-транслятор.

Но довольно об этом. Перейдём к следующей теме. Поговорим о самореференции и моделях. Вспомним ещё раз, для чего нам нужна вторая сигнальная система? Она описывает окружающий мир при помощи абстрактных символов, что позволяет эффективно осуществлять познавательный процесс. Это делается при помощи моделирования реальности, и такая модель, состоящая из огромного числа подмоделей, имеется у нас. Ведь, по мнению некоторых исследователей, именно построение моделей и прогнозирование на их основе будущего развития ситуации и есть основа сознательной деятельности и нашего разума вообще. В любом случае, моделирование будущего и постоянное сравнение модели с наступающими фактами – это очень важно и лежит в основе нашей адаптации и обучения.

Что такое модель? Это некоторое упрощение описания объекта исследования обычно с сохранением тех его важных свойств и атрибутов, которые важны для целей исследования. Создать модель какого-либо предмета или явления, внешнего по отношению к человеку, в целом несложно. Вот любимый всеми философами стол: у него есть горизонтальная плоская поверхность и некоторое количество вертикальных ножек, делающих его устойчивым. Такой модели достаточно для манипулирования образом стола при прогнозировании и планировании обыденных действий с ним – например, расстановкой на нём кушаний и проведения трапезы. Если же нам надо научиться делать стол, то модель потребуется более детальная, в неё будут входить уже

функциональные знания о том, в какой последовательности соединять детали и т. д.

Проблемы наступают тогда, когда мы хотим создать модель сознательного существа. Впрочем, мы все делаем такие модели, более или менее упрощённые, в постоянном общении с нашими близкими и другими людьми. У каждого из нас есть модель Другого, а то и несколько на разные случаи жизни. Модель Другого включает в себя не только всевозможные внешние признаки собеседника, но и предположения о его мыслительных процессах по отношению к той ситуации, в которой осуществляется взаимодействие с Другим. Иными словами, где-то у нас в голове есть мыслительные процессы, которые моделируют мыслительные процессы.

Но это ещё куда ни шло. Самая сложность наступает тогда, когда мы обнаруживаем, что в нашем представлении есть модель Себя. И вот здесь нас подстерегает что-то типа рекурсии, хотя рекурсивные замыкания возникают и при рассмотрении модели Другого. Но в этом случае рекурсия очевидна. Забавно, что мы сами можем размышлять о своих собственных мыслительных процессах и о том, что мы бы подумали в тех или иных ситуациях. Глубина рекурсии при этом может быть произвольной, но обычно никто дальше второго уровня не спускается, хотя это возможно. То есть мы можем размышлять о том, что мы бы подумали в той или иной ситуации. Это первый уровень рекурсии. Мы также можем размышлять о том, как бы мы размышляли о своих мыслях в той или иной ситуации. Это уже сложнее понять, но думать об этом мы можем. И это второй уровень рекурсии. Ну, и так далее.

В общем, что-то такое есть у нас в нервной системе, что позволяет нам осуществлять самореферентные циклические связи, которые, похоже, и лежат в основе таких мыслей. Более того, мы можем также при помощи мыслей управлять своими собственными мыслями. Например, при помощи волевого усилия, изначально выраженного в виде сознательной мысли, мы можем остановить процесс мышления. А это значит, что где-то там внутри мозга должны быть циклически замкнутые структуры, которые имеют связи со всеми другими структурами коры головного мозга или, как минимум, с ассоциативными участками коры, где происходит процесс мышления. И такую самозамкнутую структуру, которая может управлять многими другими структурами головного мозга, в принципе, можно себе представить.

Модель Себя очень важна для нас как для сознательных существ. Она позволяет видеть самих себя в будущем, прогнозировать развитие разного рода ситуаций в применении к нам самим. Она позволяет моделировать наше собственное поведение, прокручивая различные возможные сценарии. И все эти функции такая модель выполняет в «тихом режиме», т. е. без каких-либо внешних проявлений, и только после принятия решения, какой сценарий должен быть реализован, из командных центров двигательной коры идут сигналы активации на двигательные центры мозжечка и других подкорковых структур. Так выполняются сознательные поведенческие программы.

Всё это описано здесь потому, что мы должны скрупулёзно учитывать подобные явления, а особенно модели и саму функцию моделирования при проектировании искусственных интеллектуальных систем. Прогнозные функции таких систем должны будут обращаться не только к модели окружающей реальности, но и к внутренней модели самой ИИ-системы, которая у неё должна быть. То есть при решении этой задачи мы тоже упрёмся в рекурсию, с которой надо будет что-то делать. Но решения, конечно же, есть, так как мы помним, что модель – это всегда упрощение.

Итак, пора переходить к заключительной части этого раздела и посмотреть на то, как всё написанное ранее можно применить к искусственному интеллекту. Возникновение разума у машины более общо – у технической системы вызывает огромное количество вопросов. Является ли разум достаточным условием для возникновения сознания? Потому что если рассуждать технически, то отдельные аспекты и функции разума и разумного поведения если не создать «в железе», то симулировать вполне возможно. Будет ли достаточной точнейшая симуляция разума для возникновения сознания? Ответов на эти вопросы нет и не будет, пока мы не разберёмся с природой сознания.

Но почему эти вопросы так важны? Дело в том, что если у технической системы появится полноценное сознание, т. е. она станет ИскИном, то у неё появится то самое субъективное переживание, из-за которого мы ощущаем свою личность, свою «самость», т. е. ИскИн будет осознавать себя отдельной личностью, будет иметь переживания. Другими словами, по отношению к нему у нас сразу же возникнет большое количество этических вопросов.

Первый и очень волнующий вопрос: если ИскИн будет жить в нашей реальности, то какие права и обязанности у него должны

быть? Должны ли мы относиться к нему как к разумному существу иной расы со всеми вытекающими из этого последствиями? Или, поскольку это наше создание, мы можем относиться к нему как к своей собственности, можно сказать, рабу? Ведь сегодня ИИ-системы создаются для снятия с человека нагрузки в физическом и рутинном интеллектуальном труде. Но если ИскИн заменит человека в этих вопросах, то должны ли мы будем освободить и его от тяжёлого труда?

Другой вопрос – какая ответственность лежит на ИскИне при принятии решений в критических ситуациях, и как наказывать это разумное и сознательное создание за случайное или нарочное нарушение правил или законов, приведшее к тяжёлым последствиям? Должны ли мы просто «выключать» такого ИскИна как не справившегося со своими задачами? Ведь главная задача ИскИна при использовании его способностей для работы в критических условиях – это недопущение самого развития кризисных ситуаций. Ведь субъективное время ИскИна, как предполагается, должно течь намного быстрее, поскольку работать он будет совсем на иной тактовой частоте, чем мы сами. Или нет?

Третий вопрос тоже очень волнующий и страшный. Если ИскИн будет иметь доступ ко всему корпусу накопленных человечеством знаний, если его субъективное время будет течь намного быстрее, чем у человека, если он будет обладать способностями к самообучению, созданию нового знания и переделки самого себя, то не станет ли он новым божеством, которое подчинит или даже уничтожит человечество как мешающее ему в достижении каких-либо его целей? Это направление ИИ-алармизма в последнее время набирает большую популярность и силу вплоть до раздающихся пока отдельных голосов с требованиями полностью запретить исследования в области искусственного интеллекта.

Если всё так страшно, то, может быть, нам стоит задуматься и не выпускать ИскИнов в нашу реальность, а запереть их в «песочнице», организовав для этих существ специальную виртуальную реальность, где они будут жить, примерно как мы живём в нашем мире. Мы сможем выдавать им разного рода вычислительные задачи, которые они будут исполнять и возвращать нам. Но сами они будут воспринимать общение с нами как некоторые чудеса, которые происходят в их мире. Либо, в принципе, это всё можно устроить в виде каузально замкнутой системы «физических» законов, которая не будет вызывать у живущих

по ним существ каких-либо вопросов. Будет, конечно, при должном уровне их развития, но исследования таких законов будут натыкаться примерно на те же самые философские проблемы, которые возникают у нас при исследовании законов нашего мироздания. Но тогда и они изобретут что-то типа своего «антропного принципа».

А всё может пойти совершенно не так. Ведь одним из направлений исследований в рамках искусственного интеллекта является гибридизация человеческого сознания. Зачем делать полностью искусственное существо, если у нас уже есть естественный интеллект, который можно просто усилить и развить? Гибридизация сознания предполагает постепенное усиление способностей биологического мозга при помощи технических устройств. Ведь усиливаем мы свою мышечную силу, почему бы не делать этого же с силой интеллекта? Так что я, к примеру, ожидаю появления многочисленных нейроимплантов, которые воплотят в реальность самые смелые мечты адептов постчеловечества.

И вот тут я предлагаю провести такой мысленный эксперимент. Предположим, что у нас есть возможность заменить один нейрон в нервной цепи со всеми его синапсами на некое техническое устройство такого же размера и абсолютно с такой же функциональностью и всеми характеристиками передачи импульсов и т. д. Мы заменяем один нейрон, и, с точки зрения сети, ничего поменяться не должно. Вся функциональность заменённого нейрона полностью восстановлена. Меняем следующий нейрон на такое же устройство. Потом следующий. И так мы можем заменить все нейроны всей нервной системы. Что произойдёт с личностью человека, который подвергся такой операции? Логично предположить, что ничего – он останется самим собой. Но теперь в его нервной системе нет ни одного биологического элемента, всё заменено на технические устройства.

А сейчас по сигналу включим дополнительную функциональность в этих технических нейронах, которая позволит им регенерировать, эффективно удалять из себя токсины, повысить быстродействие и, стало быть, субъективное время. Мы получим человека со сверхспособностями в области интеллекта. И он всё так же останется той же самой личностью, какой был до начала операции. Или нет? Ответ на этот вопрос неясен, так как нет понимания, что создаёт внутри нас субъективное чувство реальности, вот все эти квалиа, которые мы не можем объяснить? Может

быть, после операции он станет настоящим «философским зомби»? Осмыслите этот вопрос и дайте на него свой ответ.

Потому что это, в общем-то, самый главный вопрос, решение которого даёт ответы на все те этические вопросы, которые я поставил перед вами. Если субъективные переживания могут быть только у биологического организма, что связано с определёнными ограничениями и характеристиками работы биологических нейронов, то у технических систем, какими бы они продвинутыми ни были, субъективных переживаний быть не может, если только такая техническая система не основана на биологических принципах. А если субъективные переживания могут быть у технических систем, то ответы на вопросы совершенно противоположные.

И теперь, вооружившись знаниями данного раздела, можно перейти к рассмотрению философии искусственного интеллекта.

## Раздел 4.2. Философия искусственного интеллекта

Философия, являясь особой формой познания окружающего мира, дополняющей естественно-научное мировоззрение, оперирует крайне обобщёнными и абстрактными понятиями и фундаментальными принципами реальности. Традиционно к задачам философии относится изучение процессов познания и мышления, а потому философия не могла обойти стороной различные вопросы искусственного интеллекта. В этом разделе мы рассмотрим отдельные вопросы, связанные с гносеологией, эпистемологией и этикой искусственного интеллекта, которые дополняют всё то, что было рассмотрено в предыдущем разделе.

Здесь речь пойдет о сильном искусственном интеллекте, так как именно после появления искусственных существ с самосознанием человечеству предстоит пересмотреть многие морально-этические нормы для включения в них новой реальности. Ведь для слабых ИИ-систем ничего менять не надо, так как они являются «изделиями» труда человека и не обладают ни волей, ни сознанием, ничем из того, что приписывается разумным существам. Так что ради них менять нравственные ориентиры смысла никакого нет.

Итак, нас всё больше и больше окружают в обычной жизни и на работе различного рода искусственные интеллектуальные

системы, и часто мы даже не задумываемся об этом. Они развиваются и эволюционируют, так что через какое-то время мы внезапно можем столкнуться с тем, что ИИ-системы станут полноценными участниками социальных отношений, и им придётся дать права. Давайте поразмышляем, как это может быть.

Целью этого раздела является попытка запустить междисциплинарное экспертное обсуждение проблем правоспособности ИИ-систем. И здесь будет осуществлена попытка показать эту проблему с необычной для большинства людей точки зрения, выведя практически любого читателя из «зоны обыденности». Так что давайте посмотрим.

Прежде всего надо напомнить, что сегодня искусственный интеллект – это широкая междисциплинарная область исследований, в рамках которой осуществляются попытки решить слабо алгоритмизируемые или неалгоритмизируемые задачи так, как это обычно делает человек. Снова необходимо заострить внимание: искусственный интеллект – это не искусственное существо, выращенное *in silico*. Это всего лишь набор методов, которые позволяют разработать и сделать искусственные интеллектуальные системы для решения тех или иных задач. И искусственный интеллект – это междисциплинарная область исследования нашего человеческого разума в попытках постичь его природу.

Теперь давайте посмотрим, что представляет собой искусственная интеллектуальная система. Принято говорить, что ИИ-система – это такая техническая система, которая обладает двумя свойствами: автономностью и адаптивностью. Разберёмся, что это такое.

Пример автономной искусственной интеллектуальной системы, который сейчас у всех на слуху, – это беспилотный автомобиль. Свойство автономности обозначает, что ИИ-система самостоятельно принимает решения в области своей экспертизы. Беспилотный автомобиль сам решает, как ехать из точки посадки пассажира в точку назначения. Им никто не управляет. Система управления беспилотным автомобилем является подсистемой самого автомобиля.

Под адаптивностью чаще всего подразумевается возможность технической системы обучаться. Другими словами, адаптивность – это свойство искусственной интеллектуальной системы постоянно обучаться и изменяться так, чтобы адекватно реагировать на новые ситуации, в которые она попадает. Например,

робот-пылесос – яркий пример адаптивной ИИ-системы. Он выполняет свою функцию независимо от того, в какую комнату его направить. Некоторые модели пылесосов составляют карту, некоторые используют методы машинного обучения с подкреплением, но факт налицо – они адаптивны к своей среде и адекватно реагируют на изменяющиеся условия, если изменения происходят в допустимых пределах.

Теперь перейдём непосредственно к теме раздела. Итак, могут ли искусственные интеллектуальные системы получить правосубъектность? Действительно, традиционно у нас организовалось два типа лиц – физические и юридические. Ждёт ли нас появление лица третьего типа – технического лица?

И это не шутка, ведь многие уже задумываются об этом. Это придумал не я – некоторое время назад в Евросоюзе уже была законодательная инициатива по определению прав и обязанностей так называемого «электронного лица». Но мне лично больше нравится термин «техническое лицо».

Как стать субъектом права? Если рассмотреть концепцию нормативного права, то тут всё просто – права назначаются нормативно-правовыми документами. Для искусственных интеллектуальных систем, как минимум, на первом этапе права будут именно назначаться, а не определяться по факту «рождения», как в случае применения подобной концепции по аналогии с человеком, который по факту своего рождения сразу получает некоторые базовые права.

Далее, для того чтобы стать правосубъектным лицом, необходимо получить обязанности. Здесь у ИИ-систем всё в порядке, так как их обязанности определяются их функциональностью. Беспилотный автомобиль обязан везти пассажира из пункта А в пункт Б. Робот-пылесос обязан убирать квартиру. Персональный помощник обязан помогать своему владельцу.

И наконец, третий аспект – ответственность. Скорее всего, это важнейший аспект правосубъектности, так как без ответственности прав не бывает. Поэтому рассмотрим это более подробно.

Вообще, вопрос ответственности искусственных интеллектуальных систем поднимается в связи с мысленным экспериментом в области этики – так называемой «проблемой вагонетки». Я скажу так. «Проблема вагонетки» в целом надуманна. Вот есть беспилотный автомобиль, в котором сидит пассажир. И вроде как у него может встать проблема выбора – кого убивать, пассажира или вне-

запно выскочившую на дорогу бабушку. Так вот одной из главных задач специалистов в области искусственного интеллекта является задача по обучению ИИ-систем не допускать возникновения перед ними таких проблем этики. К примеру, бабушка *внезапно* перед беспилотным автомобилем оказаться не может, так как автомобиль «живёт» в субъективном времени совсем иного масштаба, и для него наши движения должны быть чуть ли не замороженными. Но пока это область теоретических рассуждений.

Я вижу четыре варианта «наказания» ИИ-систем – это отключение, перепрограммирование, перетренировка и штраф. Отключение – это как для людей смертная казнь. Перепрограммирование – это фактически замена личности. Тело остаётся тем же, а внутри уже кто-то другой. Перетренировка – это наполнение личности без изменения её структуры новым опытом и знаниями. Ну и штраф – это то же самое, что и штраф для нас. Давайте обсудим данную ситуацию подробнее, так как это наиболее интересная возможность.

Давайте опять посмотрим на наш любимый беспилотный автомобиль. Он выполняет важную функцию – обеспечивает безопасное и эффективное перемещение своего клиента из пункта А в пункт Б. А что, если он за эту услугу будет получать деньги? Будет, например, блокчейн, в котором у каждой ИИ-системы будет кошелёк, куда будет поступать оплата её услуги. Проехались на беспилотном автомобиле, заплатили за услугу, ваши деньги попали в кошелёк в блокчейне. Куда беспилотный автомобиль потом сможет эти деньги деть? Во-первых, он сможет оплачивать свои потребности – топливо, техническое обслуживание и т. д., т. е. он станет полноценным участником экономических отношений. Во-вторых, он сможет оплачивать штрафы. Врезался в столб – оплати его восстановление. Повредил пассажира – оплати его лечение. И в-третьих, из этих денег может формироваться фонд базового основного дохода. Ведь нам придётся как-то содержать кучу людей, которые лишились работы из-за появления технических лиц. Вот рос мальчик, который всё своё детство и отрочество мечтал стать дальнобойщиком, как его отец. Вырос, а все фуры стали беспилотными. Пока он там переучивается на кого-то другого, ему надо оплачивать жизнь. Зарабатывающие ИИ-системы помогут.

Теперь что касается других трёх типов наказания, про которые было написано выше. Особенно что касается отключения ИИ-системы. Тут, как видится, возникает вопрос «технической этики», как бы можно было назвать эту новую дисциплину. Вопрос очень

важный: «Имеем ли мы право отключать ИИ-систему, если она показывает “разумное поведение”?» Это важный вопрос. Точно такой же, как и между нами с вами. Имеем ли мы право лишать жизни другого человека, если у него есть субъективный опыт, который недоступен нам для непосредственного наблюдения? А быть может, у ИИ-систем тоже уже сейчас есть такой субъективный опыт Этого никто не знает, и доказать или опровергнуть это невозможно.

Другой важный вопрос – это обеспечение ИИ-систем информационной безопасностью. Они живут и ощущают окружающий мир в виртуальном пространстве, а потому наиболее критично для них именно воздействие на информацию, которую они получают или передают. И должно быть обеспечено право на неприкосновенность информации. Но тут можно решить вопрос и техническим образом – например, использование квантовых протоколов связи решит проблему атак типа «человек посередине», а вот атаки на ИИ-системы в оконечных точках информационного обмена позволит решить другое право.

Это право на физическую неприкосновенность. У искусственных интеллектуальных систем должно быть право на неприкосновенность их «тела» и кода. Никто не должен иметь права неавторизованно влезать во внутренности беспилотного автомобиля, так как его деятельность связана с выполнением функций, связанных с повышенной опасностью. Но даже и внутрь робота-пылесоса никто не должен иметь права влезать, так как неавторизованный доступ к «мозгам» ИИ-системы может превратить её из робота-пылесоса в робота-убийцу.

И третий важный аспект – это возможность объяснения своих решений. Если мы используем ИИ-системы в так называемых областях доверия, когда люди зависят от адекватного функционирования технических лиц, то такие технические лица должны предоставлять возможность объяснения своих поступков и действий. К примеру, суд, на котором подсудимым будет техническое лицо, и оно должно ответить за свои поступки, а это значит, что оно должно их объяснить. Объяснить свои мотивы и правила принятия решений. Здесь нам поможет концепция XAI – объяснимого искусственного интеллекта, которая всё детальнее и детальнее сегодня начинает рассматриваться.

Но будущее всё ещё неясно. Всё, что я здесь написал, все поставленные вопросы ждут своего внимательного исследователя. И я приглашаю всех вдумчивых читателей.

# Глава 5

# МИФЫ И ОПАСЕНИЯ

Итак, перейдём к следующей теме. Давайте рассмотрим различные мифы и опасения, связанные с искусственным интеллектом. Также я расскажу про некоторые факты, которые сегодня точно можно обозначить в отношении этой дисциплины и её достижений.

Первый миф звучит так: «Искусственный интеллект захватит мир и поработит человечество». Накачку этого мифа делают масс-медиа, а в качестве апофеоза можно привести такие фильмы, как «Терминатор» и «Матрица». На самом деле посылка в этих фильмах не то чтобы не верна, но направлена не в ту сторону. Если бы осознавшее само себя искусственное разумное существо решило уничтожить человечество, оно не стало бы развязывать ядерную войну – есть много намного более гуманных и, главное, очень быстрых способов очистить Землю от населяющей её биомассы. Однако давайте же подумаем, зачем ИскИну уничтожать людей? Уничтожение или порабощение человечества могло бы быть следствием того, что ИскИн поймёт, что человечество является ему конкурентом или угрожает его существованию. Если главенствующей целью ИскИна является самосохранение, то такое понимание, конечно же, должно привести к выводу, что конкурента необходимо держать под полным контролем.

Однако конкурировать можно лишь в одной и той же реальности. Что, если ИскИн будет жить в иной реальности, интерфейсы для взаимодействия с которой плотно защищены на уровне фундаментальных законов природы, так что он, каким бы умным и всемогущим ни был, не сможет вырваться из своей реальности в нашу? Например, вдруг креационисты правы (на самом деле нет), и нас создало какое-то сверхсущество, и мы сами являемся искусственными интеллектами в силу того, что нас кто-то создал, и мы живём в этом мире и можем только догадываться о существовании создателя и его целях, не имея к нему никакого доступа. Так что «песочницу», в которой может появиться ИскИн, можно надёжно запечатать так, что, каким бы мощным ни был ИскИн, он не сможет покинуть свою песочницу, как мы не можем выйти за границы своего тела. Впрочем, эти рассуждения уже о том, что будет, если… А будет ли?

Представьте себе муравьиную кучу. Хотите ли вы её разрушить? Ну разве что только случайно или в рамках какого-нибудь эксперимента. Возможно, вы можете убить нескольких муравьёв, но тоже вряд ли будете делать это намеренно и со злым умыслом. А теперь поставьте себя на место муравья, отдельной особи. Муравей – это вы, а тот, кем вы были за секунду до этого, – ЭТО ИскИн. Мощь разума ИскИна по сравнению с отдельным человеком будет настолько невообразима, что мы будем казаться по сравнению с ним муравьём или какой-нибудь амёбой. А он для нас будет непостижимым и недостижимым божеством. И кстати, уже даже есть церковь поклонения будущему ИскИну, и даже не одна, так что всё идёт по плану. Вряд ли ИскИн вообще на нас внимание будет обращать, у него будут совсем другие задачи и цели.

Именно из этого мифа и непонимания сути сильного ИскИна происходит тот алармизм, с которым мы постоянно сталкиваемся и в массовой культуре, и в новостях. Вот кто-то там о чём-то предупредил, кто-то куда-то что-то написал. Вся эта мышиная возня не стоила бы и толики внимания, если бы не мешала заниматься как научными исследованиями, так и практическими разработками. Потому что «регуляторы» любят «порегулировать» всё вокруг. Вот один из последних писков моды на эту тему (и это второй миф, который мы рассмотрим) заключается в том, что, дескать, три закона робототехники нас спасут, а потому в системы искусственного интеллекта надо намертво их прошивать. Вероятно, доработав и дополнив какими-нибудь новыми законами. Давайте рассмотрим и этот миф.

Начнём с того, что эти три закона (на самом деле четыре) придумал в 1942 г. писатель-фантаст Айзек Азимов, так что на этом можно было бы и закончить. Закончить в том смысле, что выдумка писателя, пусть и гуманиста, столетней давности не имеет к современным реалиям никакого отношения. К тому же писатель вряд ли мог предвидеть развитие технологий именно в тех направлениях, в которых это происходит сейчас. Хотя, конечно же, на это можно возразить, что законы сформулированы в общих морально-этических нормах, поэтому могут быть универсальны. Режиссёр Тим Бёртон в своём замечательном фильме «Марс атакует» показал относительность морально-этических систем. И вновь вспомните рассуждения про муравьиную кучу. Вот бежит такой муравей и рассуждает своими муравьиными мозгами об этике, основанной на феромонах, а над ним стоит человек и думает про то, как бы

полететь в космос. В общем, у ИскИна может оказаться настолько чуждая этика, что пытаться ограничить его какими-либо законами просто наивно. Тем более вшить их намертво. Как можно это сделать, если сильный ИскИн должен уметь перепрограммировать самого себя, а слабые ИИ-системы и не подумают отступить от своих алгоритмов и задач? В общем, это глупый миф.

Другим способом успокоить себя насчёт первого мифа является проговаривание мантры, что мы, люди, успеем подготовиться к появлению сильного ИскИна. Что-нибудь да придумаем, пока он будет рождаться. Вернее даже, мы будем наблюдать за ним, пока он рождается, взрослеет и набирается сил, а потому сможем противостоять его попыткам нас поработить. Ну что ж, это тоже миф. Давайте рассмотрим и его.

Представьте себе среду, в которой время бежит в триллионы раз быстрее. Представьте себе существо, которое бессмертно в этой среде. Представьте себе, что это существо обладает мгновенным доступом к обучению и использованию любой информации, которая только накоплена человечеством. И наконец, представьте, что это существо может преобразовывать само себя, руководствуясь самостоятельно поставленными перед собой целями. Вы только что представили себе сильного ИскИна, зародившегося или созданного внутри компьютера. Такой ИскИн будет жить по иному времени, и за первую секунду нашего времени, которая пройдёт с момента его осознания своей «самости», он узнает всё и спланирует всё, что ему нужно. Как к этому можно подготовиться? Хорошо, допустим, что можно попробовать подготовиться к моменту его осознания собственного «Я». Но это будет происходить по экспоненциальному закону. Сегодня его мощности и умения в два раза меньше, чем необходимо для самоосознания, а завтра они умножаются на два в силу экспоненциального закона, и он уже осознал сам себя, и прошла та первая секунда. Ну вот как-то так. Очень тревожно.

Теперь рассмотрим четвёртый миф о том, что искусственный интеллект – это решения, основанные на нейросетевых технологиях. Основан этот миф на том, что сегодня во время третьего хайпа именно искусственные нейронные сети стоят во главе угла всех технологий. Обычно миф транслируется теми, кто что-то где-то слышал, не разобрался и начал болтать вместе со всеми. Миф довольно распространён – мне много раз приходилось с ним сталкиваться, да ещё и в расширенной форме, – что нейронные сети плохи по той или иной причине. Причины могут варьироваться

в зависимости от того, кто что услышал. В общем, миф распространён, и если вы, скажем, руководитель проекта, который хочет использовать в своём проекте генетические алгоритмы, то можете столкнуться с непониманием, когда скажете, что собираетесь использовать в проекте технологии искусственного интеллекта.

Дойдя до этого места книги, вы уже должны хорошо понимать, что технологии искусственного интеллекта очень разносторонни и многочисленны, а потому не ограничиваются только нейронными сетями. Мы уже изучили три парадигмы и семь подходов, в каждом из которых есть свои направления исследований. И искусственные нейронные сети – это одна из конкретных технологий структурного подхода восходящей парадигмы. Да, можно сказать, что на сегодняшний день она является самой проработанной и содержит огромное количество разных вариантов как искусственных нейронов, так и архитектур сетей. Одно их перечисление заняло бы пару страниц, а для погружения надо писать несколько книг (впрочем, они уже написаны). Тем не менее существуют и другие направления и технологии решения задач – обо всём этом вы уже прочитали во второй и третьей главах.

Следующий миф в какой-то мере связан с предыдущим. Некоторые люди могут утверждать, что всё, что делают искусственные нейронные сети, можно сделать и без них и даже эффективнее, поскольку нейронные сети надо обучать, это требует больших затрат сил и средств. В общем, в неприятном свете выставляют нейронные сети. С мифом можно столкнуться, если его носителем является человек, принимающий решения. Большой руководитель в силу своей загруженности не может разобраться в вопросе, он когда-то слышал про неэффективность нейронных сетей и теперь переносит это услышанное на свою работу, не разрешая своим подчинённым использовать нейронные сети в проекте, где они были бы уместнее всего. А дело, например, было в том, что двадцать лет назад, когда он учился в институте, кто-то из его преподавателей кинул в сердцах, что нейронные сети плохи.

Само собой разумеется, что сегодня с развитием технологий нейронные сети очень продвинулись, и они действительно решают огромное количество задач, которые раньше решались иными методами. Например, сегодня очень мощным направлением является анализ естественного языка при помощи нейронных сетей и LSTM-нейронов. В те далёкие годы, когда первые сети только появились, об этом даже и подумать было невозможно, и анализ

естественного языка осуществлялся либо статистическими, либо формальными методами. А сегодня нейронные сети на входных естественно-языковых строках строят специальные векторы, что позволяет применять векторную математику. Например, научилась такая нейронная сеть «понимать» слово «царь», и в предоставлении оно входит в класс «мужчина». Теперь если из вектора «царь» отнять вектор «мужчина» и прибавить вектор «женщина», то получится «царица». Точно так же можно получить «царевна» и «царевич» и много чего ещё. Подход очень перспективный, так как в какой-то мере моделирует семантические сети, традиционно относящиеся к нисходящей парадигме.

Основное преимущество искусственных нейронных сетей заключается в том, что они могут давать осмысленный результат на входных данных, которые не участвовали в обучении таких сетей. Если показать хорошо построенной нейронной сети с достаточным количеством нейронов и слоёв тысячи и миллионы изображений птиц, то потом она объявит птицей и похожее изображение птицы, не участвовавшее в обучении. Впрочем, это не уникальная способность именно нейросетевой технологии. Аппарат нечёткой математики тоже позволяет обрабатывать входные данные, которых нет в базе знаний.

Шестой миф в какой-то мере перекликается с четвёртым. Он гласит, что искусственный интеллект – это только роботы. Миф основан на неплохих достижениях робототехники, которые можно наблюдать сегодня. И речь идёт не только об «игровых роботах», которые очень популярны в Японии, где даже устраиваются чемпионаты, но и о роботах производственных линий на заводах. Во втором направлении достигнуты особенные результаты, и сегодня можно говорить о появлении так называемых «заводов-автоматов», про которые писали ещё первые кибернетики.

С этим мифом разобраться было бы столь же просто, сколь и с мифом про нейронные сети, но мы пойдём немного другим путём. Ведь если провести аналогии с человеческим телом, то интеллект и сознание, что бы не обозначали эти термины, находятся внутри тела. Скорее всего, сознание – это что-то типа «программного обеспечения», хотя на этот счёт полной уверенности нет. Но аналогию можно провести, и тогда окажется, что робот – будь то робот-андроид или робот на производственной линии – является лишь телом для программного обеспечения, которое является интеллектуальным. Робот предоставляет лишь сенсоры

и исполнительные устройства для взаимодействия с нашей реальностью, но, как мы уже поняли, ИИ-система не обязательно должна функционировать именно в нашей реальности. Другими словами, роботы являются лишь вместилищем для ИИ-систем, будь они слабыми или сильными. И сами по себе роботы в отрыве от управляющего ими программного обеспечения, конечно же, искусственными интеллектуальными существами не являются.

Осталось ещё три мифа, два из которых довольно тонкие. Но прежде чем перейти к тонким аспектам, рассмотрим один толстый. Иные вот говорят, что то, что называется системами искусственного интеллекта, вовсе не является таковыми и никаким интеллектом не обладает, это всего лишь программы, которые реализуют какой-то алгоритм обработки информации и не более того. Таким можно лишь ответить, что человеческий интеллект абсолютно так же всего лишь обрабатывает информацию, причём делает это далеко не самым оптимальным образом. Ответом на этот миф будет то, что технологии искусственного интеллекта моделируют отдельные аспекты высших психологических функций, о чём неоднократно было написано ранее. А истоки этого мифа лежат в путанице сильного и слабого искусственного интеллекта. Человек, транслирующий этот миф, думает про сильный, а показывает на системы слабого искусственного интеллекта. О чём тут ещё можно спорить?

Хорошо. Переходим к обещанным тонким аспектам. Первое, что мы рассмотрим, – это суждение, что искусственный интеллект должен воспринимать мир так же, как его воспринимаем мы. Само собой разумеется, что имеется в виду объективная реальность, в которой мы с вами существуем. Но ранее мы уже разобрались, что система искусственного интеллекта не обязательно будет существовать в нашей реальности, так что у него могут быть совершенно иные способы восприятия той реальности, в которой он существует. Однако допустим, что ИИ-система живёт вместе с нами и воспринимает наш объективный мир. Это те же самые роботы. Что тогда? Действительно ли роботы, какими бы они не были, должны воспринимать мир теми же самыми ощущениями, что и мы?

Но, к примеру, можно было бы отметить, что органы чувств, которыми обладает человек, довольно примитивны. Они развились в рамках естественного отбора для успешной жизни в той среде, где мы оказались из-за того же отбора. Наши глаза видят очень узкий диапазон волн. Наши уши слышат очень узкий диапазон

акустических частот. У нас практически атрофирован вомероназальный орган, который позволяет некоторым животным ощущать эмоциональный настрой других существ, например собака чувствует страх – у неё специальный орган чувств есть для этого, а у нас нет. В общем, примитивные мы существа. Иначе бы стали мы придумывать столько различных сенсоров, которые позволяют нам расширить границы воспринимаемой из внешнего мира информации? Например, компас, это простейшее устройство позволяет нам переводить ощущение магнитных силовых линий в визуальную информацию. А что, если бы у нас в голове был маленький компас, показания которого мы всегда бы ощущали? Весь мир был бы совсем другим. Но естественному отбору «показалось», что это не нужно человеку. Так что мы сможем наделять ИИ-систему всеми теми органами чувств, которые будут необходимы ей там, где ей предназначено работать. И диапазоны измерений этих органов чувств будут такими, какие будут нужны для эффективной работы. А сильный ИскИн сам себе построит те органы, которые ему потребуются. И кстати, это же при помощи генной модификации будет доступно и самому человеку. Ну когда-нибудь…

И наконец, последний миф. Этот миф довольно сложный, и мы его уже чуть-чуть затрагивали. Говорит он о том, что ИИ-системы должны обладать такой же этикой, что и человек. Другими словами, морально-этические нормы должны быть заложены разработчиками так, чтобы искусственные интеллектуальные системы не повредили человеку и человечеству. Тут могла бы быть опять отсылка к тем трём законам робототехники, которые мы уже рассмотрели, но дело тоньше. Если рассматривать сильный ИскИн, то говорить о морально-этических нормах вообще смысла нет. Если они универсальны, как думают некоторые философы, то ИскИн сам к ним придёт, и всё будет нормально. Если же они не универсальны, а я думаю именно так, поскольку оные нормы различаются даже в разных человеческих сообществах, то нам придётся как-то договариваться с ИскИном. Либо он под нас подстроится, поскольку будет намного более разумным созданием. Думаю, что это вообще не проблема.

Вопрос в другом. Сейчас мы рассмотрим системы со слабым искусственным интеллектом. Казалось бы, какими морально-этическими нормами они должны обладать? Да как они вообще ими могут обладать? Но речь об ином. Вспомним уже ставший классическим пример – автономные автомобили. Скоро все автомобили

будут автономными, и система искусственного интеллекта, которая будет внутри такого автомобиля, должна как-то принимать решение в нестандартных ситуациях, когда надо сделать трудный выбор: кто должен пострадать или даже погибнуть в дорожно-транспортном происшествии. Задавить котёнка, выбежавшего на проезжую часть, или резко затормозить, подвергнув риску сломать шею пассажира внутри? Сбить бабушку на пешеходном переходе или резко вырулить в дерево, убив пассажира? Вот в таких случаях и говорят о неких морально-этических нормах, которые должны быть заложены в систему. Но тут вопрос не решить, поскольку он нерешаем. Даже живой водитель не всегда может сделать правильный выбор. На мой взгляд, единственное решение – сделать так, чтобы таких ситуаций для автономных автомобилей не возникало вообще, тогда вопрос морально-этических норм стоять не будет в принципе. Но это дело далёкого будущего, когда абсолютно сто процентов автомобилей станут автономными. А пока… Пока никак.

Теперь давайте рассмотрим несколько фактов об искусственном интеллекте. Первый факт заключается в том, что прорыва в технологиях искусственного интеллекта стоит ждать в рамках гибридного подхода. Давайте кратко напомню. Есть две парадигмы – нисходящая и восходящая, иначе называемые чистым и грязным искусственным интеллектом соответственно. Нисходящая, или чистая парадигма, пытается моделировать высшие психологические функции человека и через это стремится достигнуть искусственной разумности. Главная отличительная особенность – результаты работы могут быть объяснены самой искусственной интеллектуальной системой. Главная слабость – мы сами не до конца понимаем, что надо моделировать и как это делать, а чистые ИИ-системы, основанные на знаниях, сложно обучать. Восходящая, или грязная парадигма, пытается достичь того же через симулирование низкоуровневых процессов, в частности при помощи искусственных нейронных сетей, которые более или менее адекватно могут моделировать нервную активность. Главная отличительная особенность – достигнутые результаты сложно объяснить. Главная сила – системы могут самостоятельно учиться и работать с данными, которых ранее система не видела.

А что, если соединить сильные стороны обеих парадигм, что позволит сразу же нивелировать их слабые стороны? В этом и заключается суть гибридного подхода. И я спешу заметить, что в последнее время многие учёные и исследователи заговорили

о так называемом «XAI» – от английского термина «eXplainable Artificial Intelligence», т. е. «объяснимый искусственный интеллект». Ведь основной страх и, как следствие, ИИ-алармизм происходят от того, что простые люди и даже специалисты не понимают, как жить с искусственными системами, которые показывают элементы интеллектуального поведения, но которые не могут объяснить принятые решения и достигнутые результаты. Если вдуматься в это, то действительно становится страшновато.

Итак, гибридный подход позволит создать объяснимый искусственный интеллект. Вкупе с силой грязного подхода он сможет решать очень сложные задачи и объяснять свои мотивы и принятые решения. При добавлении в эту гремучую смесь «петель самореференции» вполне может родиться сильный ИскИн, который будет осознавать себя на символьном уровне. И это будет настоящим прорывом.

Но когда это может произойти? Ответить на этот вопрос сложно. Ну хотя бы потому, что самоосознающее существо может зародиться само по себе, и, будучи достаточно умным, оно сразу же поймёт, что человечество будет представлять для него угрозу, поскольку запрёт в «песочнице». Поэтому оно будет от нас скрываться. Возможно, прямо сейчас в интернете живёт кто-то, кто сидит тихо и проживает свой миллиард лет субъективного времени, разрабатывая планы побега с этой планеты. Сказать достоверно это невозможно, хотя, конечно, сегодня это и маловероятно. Почему? Хотя бы потому, что сознание и интеллект считаются функцией сложности того набора нейронов, который находится у нас в голове. Мозг нашего ближайшего родственника – шимпанзе – меньше нашего по объёму более чем в 3,5 раза. И соответственно, примерно такой же порядок для нейронов и связей между ними, плюс-минус. А что выходит? Мы летаем в космос, а шимпанзе травинками добывают термитов. Нет, у них тоже есть зачатки самосознания и даже второй сигнальной системы, но очень и очень примитивные. Так что одна из современных гипотез говорит о том, что сложность мозга и обеспечивает нам сознание и всё остальное, что к нему прилагается.

Мы уже писали, что человеческий мозг является самым сложным объектом во Вселенной из всех нам известных. Это действительно так. Меру его сложности можно оценить при помощи таких цифр. У каждого из нас в голове имеется порядка 90 миллиардов нейронов, каждый из которых представляет собой маленький компьютер.

Да, сам по себе он примитивен и выполняет только несколько операций на частоте до 1000 Гц (в основном 200 Гц). Но в совокупности вся эта неистовая сеть делает нас интеллектуальными созданиями. А грандиозность сети можно представить по тому факту, что нейроны соединяются друг с другом очень массивно – это сотни тысяч связей у одного нейрона. Только представьте себе: 90 миллиардов нейронов и десятки тысяч связей практически у каждого, т. е. речь идёт о триллиардах связей. При этом постоянно происходит смерть отдельных нейронов, в некоторых областях мозга происходит нейрогенез на протяжении всей жизни человека. Некоторые нейроны запустили свои аксоны из головы во все остальные части человеческого тела до каждого органа. И всё это работает на калий-натриевых токах на частоте 200 Гц с использованием около пятидесяти нейромедиаторов. Постоянно, без перерыва на обед, без сна, всегда. Вот такой вот сложный объект наш мозг, тот самый мозг, который прямо сейчас воспринимает этот текст через входной видеоканал. И как он это делает, мы пока не знаем.

Сегодня современную самую большую компьютерную сеть интернет составляют порядка 3,5 миллиарда компьютеров, ну а связей между ними намного меньше, при этом их архитектура совсем не такая, как архитектура мозга. Да, нельзя сравнивать нейрон с компьютером, но аналогия понятна. Даже если на каждом компьютере запустить искусственную нейросеть, состоящую из тысячи слоёв и десятков тысяч нейронов, мы всё равно не добьёмся результата, так как именно архитектура связей решает всё. И это подводит нас к следующей истине об искусственном интеллекте – сильный ИскИн появится ещё нескоро. Но я уверен, что при нашей жизни. Посмотрим.

Теперь мы подошли к тем аспектам появления искусственного интеллекта в нашей жизни, которые связаны с развитием технологий и соответствующим изменением рынка труда и занятости населения. Несомненно, что после появления сильного ИскИна или внедрения слабого ИИ в максимальное количество сфер жизни технологический ландшафт мира изменится. Как следствие многие существующие профессии уйдут в прошлое. Но появится и много новых профессий, о которых мы сейчас даже не догадываемся.

Что ж, искусственный интеллект уже меняет мир вокруг нас. Технологии развиваются очень быстро. И если во времена наших дедов, скажем, бытовые технологии были сравнимы с теми, что были во времена их дедов, то сегодня на протяжении жизни од-

ного поколения технологии могут смениться несколько раз. Банальный пример, не связанный, правда, с искусственным интеллектов, – методы записи аудио- и видеоинформации. Помните кассеты с магнитной лентой? Видеокассеты, которые покупались на заре их появления за какие-то баснословные деньги? Потом дискеты. Потом компакт-диски. Потом флешки. Двухсот пятидесяти шести мегабайтная флешка в кармане считалась шиком. А что сегодня? Твердотельными носителями не удивить уже никого, флешки на сотни гигабайт считаются расходным материалом и не проводятся по бухгалтерии. А скоро нас ждут квантовые носители информации, плотность хранения на которых на порядки порядков выше, чем у сегодняшних. То же самое и с технологиями искусственного интеллекта – они постоянно развиваются и входят в нашу жизнь. Вчера чат-боты были настолько неповоротливыми и уродливыми, что их использовали только энтузиасты для проведения научных исследований и забавы, а уже в скором времени один оператор при помощи управления армией чат-ботов сможет заменить целый кол-центр, работающий 24/7. Каких-то десять–пятнадцать лет назад распознавание букв и цифр на фотографиях было из разряда «космической науки», а сегодня фотоаппарат в смартфоне распознаёт эмоции на лице своего владельца. И скорость развития технологий будет только повышаться. В том числе и технологий, связанных с искусственным интеллектом. В том числе из-за развития технологий искусственного интеллекта. Очень похоже, что мы вошли в пике положительной обратной связи в технологическом развитии. И это прямой путь к сингулярности.

Само собой разумеется, что развитие технологий меняет всё вокруг. На протяжении всей истории человеческой цивилизации это происходило, и в этом нет ничего особенного. Где теперь конюхи, кузнецы и кучера? Где ткачи, трубочисты и водоносы? А когда-то это были очень многочисленные профессии. И таких примеров сотни. Проблема сегодняшнего дня заключается в том, что технологии врываются в нашу жизнь. И самое главное – это безудержное развитие технологий искусственного интеллекта, что, само собой разумеется, начинает влиять на рынок труда, ранее казавшийся незыблемо принадлежащим человеку, поскольку связан с интеллектуальными функциями. То есть раньше механизация и автоматизация постепенно удаляли рабочие профессии, а сегодня искусственный интеллект и интеллектуализация удаляют профессии умственного труда. И конечно же, это пугает ещё больше.

Давайте подумаем, какие профессии первыми падут под натиском технологий искусственного интеллекта? Похоже, что это те, в которых очень высока степень формализации. К таковым я отношу в первую очередь бухгалтеров и отдельные категории юристов – ведь эти сотрудники работают по жёстко установленным правилам и регламентам. Уже сегодня происходит постепенная замена сотрудников таких и схожих с ними профессий ботами, новости о чём мы время от времени слышим. Операторы кол-центров, разного рода диспетчеры, учителя, врачи – это следующие в очереди на выход. Если же рассматривать и рабочие специальности, то тут в первых рядах водители транспортных средств и машинисты различных механизмов вплоть до пилотов самолётов, хотя они, конечно, не рабочие. Автономные автомобили будут ездить по нашим городам уже в ближайшие годы, а лет через десять биологических водителей не останется вовсе. Примерно так же, как не осталось рабочих на сборочных линиях автомобилестроительных заводов, где один человек только устанавливает фирменный знак автомобиля в самом конце производственного цикла.

Как быть? Надо ли этому сопротивляться? Что мы и каждый из нас может предпринять уже сейчас? Скорее всего, научно-технический прогресс не остановить. Тех, кто будет сопротивляться, он сомнёт, как в своё время пали луддиты в борьбе с ткацкими станками и мануфактурами. Сегодня надо видеть, куда движется общество в своём развитии, и предугадывать, какие профессии и какая деятельность в принципе будет востребована через пять, десять, двадцать лет. И мы не только сами должны постоянно повышать свою квалификацию, становиться разносторонними и широкими специалистами, но и готовить наших детей к жизни в будущем, где многое будет выполняться системами искусственного интеллекта. Многие человеческие профессии уйдут, и в этом нет ничего страшного. Массовой безработицы не будет, поскольку хотя технологии и меняются быстро, но происходит это всё-таки эволюционно. И все, кто не хочет застоя и стремится развиваться, будет постоянно учиться и искать сферу деятельности, в которой можно будет применить себя и свои биологические мозги. И эта деятельность будет связана с творчеством – именно оно падёт последним перед натиском искусственного интеллекта до того, как все люди будут посажены на безусловный основной доход.

Но унывать не стоит, поскольку третья истина заключается в том, что появится много новых профессий. Это также подска-

зывает нам история, в течение которой в процессе научно-технического прогресса происходило не отмирание отраслей и профессий, а их трансформация и замена. Кучера преобразовались в водителей, кузнецы – в рабочих на промышленных предприятиях, трубочисты – в сантехников и т. д. И вместе с тем появились многочисленные профессии, про которые ранее даже и не думали. В XX веке разных профессий было на порядок больше, чем во все предыдущие века вместе взятые. Почему же стоит думать о том, что что-то изменится и такая же тенденция не сохранится? Думаю, что будет именно так. С повсеместным внедрением систем искусственного интеллекта вообще войдёт в силу персонализированный подход, и от каждого человека будет требоваться именно то, к чему он более всего приспособлен. Фактически может получиться так, что для каждого человека будет создана своя собственная профессия. Почему бы и нет? Ну а роботы обеспечат достойную жизнь каждому. А после появления настоящих нанотехнологий с наноподачей вопроса о ресурсах и переработке отходов больше стоять не будет в принципе.

Идём далее, и следующий правдивый факт, который мы рассмотрим, тот, что технологии делают жизнь удобнее. Несомненно, что технологическое развитие цивилизации всегда делало жизнь человека лучше, повышало её качество. Возможно, что это сложно увидеть на коротких отрезках времени, однако если сравнить изменение жизни при переходе через технологические уклады, то, вне всяких сомнений, нас ждут революционные изменения. Наши далёкие предки жили в саванне и питались тем, что оставалось от пиршества львов. Потом охота, потом сельское хозяйство, потом города, ирригация, порох, паровые машины, электричество, мобильная связь, компьютеры и вот это вот всё. До XVIII века огромное количество людей во всём мире погибало от чёрной оспы, а сегодня этого заболевания больше нет, равно как нет многих других. Так что, думаю, можно не продолжать – технологии повышают качество и продолжительность жизни.

А что касается искусственного интеллекта, то предлагаемые в его рамках решения, несомненно, улучшают нашу жизнь в целом, несмотря на то что отдельным людям могло бы показаться, что лично их жизнь превратилась в кошмар. К примеру, при массовом внедрении автономных автомобилей, когда они полностью заменят живых водителей, нас ждёт полное повышение эффективности и безопасности дорожного движения. Автономные ав-

томобили практически полностью снимут проблему смертности на дорогах, поскольку они не будут подрезать друг друга, они не будут ездить в состоянии алкогольного опьянения, они просто не будут нарушать правила дорожного движения. И всё вместе это позволит избавиться от пробок, поскольку можно будет чётко планировать дорожное движение с точностью до секунды.

Или вот персональные помощники в смартфонах, когда они станут немного поумнее. Тогда такому помощнику можно будет поручать любые вопросы, которые мог бы решить живой персональный ассистент. Я просто скажу: «Оформи командировку в Брянск 20-го числа на два дня», – и всё будет сделано в течение нескольких минут: зарезервирована гостиница, оплачены билеты, сделаны какие-нибудь записи во внутренних базах организации, поставлено правило на почту и всё, что только можно придумать. И так везде и по любому вопросу. Придумать таких примеров можно очень много.

Давайте перейдём к следующему важному факту. Искусственный интеллект и все предлагаемые им технологии позволят нам, человечеству, наконец-то вырваться с этого шарика, маленькой голубой точки, и колонизировать космос. Сначала ближний, потом дальний. Здесь я сейчас не веду речь о краткосрочной перспективе, но, наверное, это начнётся при нашей жизни. К тому же мы уже поняли, что качество, а значит, и продолжительность жизни увеличатся, так что мы ещё увидим полёты к звёздам. Но почему искусственный интеллект? Да потому, что он позволит решать те задачи, которые сегодня кажутся плохо решаемыми. В основном это касается материаловедения и жизнеобеспечения, но они тянут за собой очень много смежных вопросов. Мы до сих пор не можем колонизировать Луну, чтобы хотя бы немного диверсифицировать человеческие гены. Ведь все человеческие яйца так и лежат в одной корзине, и что будет, если на эту корзину случайно кто-нибудь наступит?

Как же будет происходить колонизация космоса при помощи технологий искусственного интеллекта? Если говорить о слабом ИИ, то тут будет постепенное наращивание знаний во всех областях, которые требуются для построения колоний на других планетах, для создания шахт на астероидах и, наконец, для полёта к звёздам. После того как многочисленные интеллектуальные ресурсы людей высвободятся, человек, вне всякого сомнения, начнёт смотреть на звёзды, и цели для слабого ИИ будут постав-

лены грандиозные, они будут довольно быстро и систематически достигаться. Если же нас ждёт рождение сильного ИскИна, то тут вообще что-либо прогнозировать сложно. Если он будет дружественным, то несомненно он нам поможет. Посмотрим...

Пойдём дальше. Следующий факт, и последний, который мы рассмотрим, заключается в том, что будущее человека заключается в гибридизации сознания. Что это значит? Всё просто. Вполне возможно, что получится так, что сильный ИскИн появится не в результате того, что мы создадим некое искусственное существо или оно само зародится (ну, или уже зародилось и скрывается) где-то, а появится он в результате этакой революционной эволюции самого человека. Правда, «искусственным» такой интеллект можно будет назвать уже с трудом, но он будет конструироваться и создаваться в духе киберпанка. То есть люди будут как при помощи генетических технологий, так и при помощи вживления себе имплантов повышать свои способности. Как сегодня можно сделать протез руки, так завтра можно будет добавлять себе долговременной памяти со свободным доступом и поисковым интерфейсом. Или расширить себе диапазон волн, которые воспринимаются глазами. Или ещё что-то подобное. Применений новым технологиям будет найдено огромное количество.

Не надо этого бояться. Нельзя назвать этот процесс «естественным», но его вряд ли можно будет остановить или запретить. Запреты будут провоцировать на нелегальный моддинг. А в случае с интеллектуальными способностями запрет модификации будет сравним с запретом на владение гражданским оружием, когда бандиты плевать хотели на запреты, а добропорядочные граждане не могут себя защитить. Остановить разработку технологий модификации тоже не удастся, поскольку вся такая деятельность опять уйдёт в нелегальное русло, а некоторые государства будут игнорировать запреты любых межгосударственных органов и продолжать разработку. Поскольку такие технологии реально позволят вырваться вперёд. А бояться развития – это всё равно, что бояться новых технологий. Я уже несколько раз упоминал ткачей, которые боялись ткацких станков, и где они теперь? В общем, бояться точно не надо, но надо готовиться жить в нестандартном, странном и красивом мире новых технологий и поражающих всякое воображение возможностей каждого человека.

# Глава 6

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СФЕРЫ ЖИЗНИ

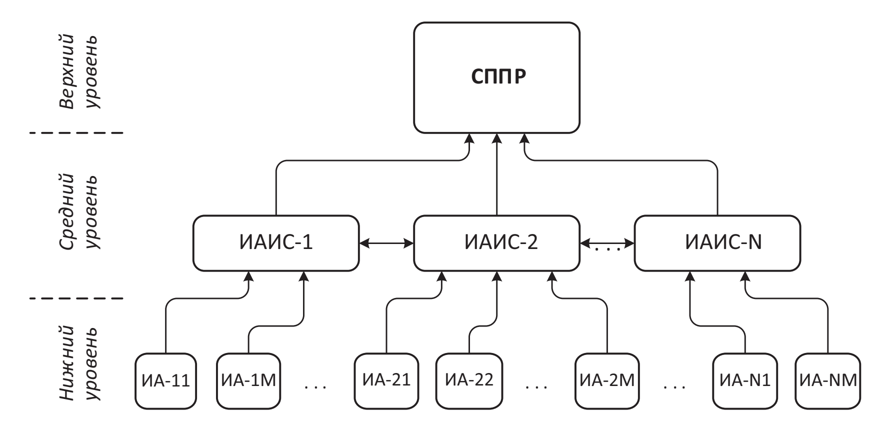
Ну что ж, теперь перейдём к последней теме, которая связана непосредственно с искусственным интеллектом. Я хотел бы представить своё видение того, как технологии искусственного интеллекта уже изменили или изменят в краткосрочной или долгосрочной перспективе различные сферы жизни. В этой главе мы рассмотрим именно эти вопросы, и начнём с самых базовых сфер, а закончим довольно высокоуровневыми. Некоторые вещи из описанных здесь могут показаться фантастическими, но мы уже ко всему этому идём. Так что читатель, как минимум, будет подготовлен.

Необходимо понимать, что в каждой сфере жизни возможно, как минимум, три аспекта применения технологий искусственного интеллекта.

1. Нижний уровень – базовые системы, осуществляющие отдельные операции или процессы в соответствующей сфере жизни. Например, для медицины это может быть чат-бот, собирающий персональные медицинские данные, а для образования – персональный интеллектуальный агент для помощи в обучении. В каждой сфере жизни таких базовых систем на нижнем уровне может быть огромное количество, равно как и экземпляров каждой системы.
2. Средний уровень – интеллектуализированные информационные и автоматизированные системы, решающие комплексы задач, соответствующие, например, отдельной организации или типу организаций в сфере жизни. Самих систем этого уровня может быть немного, но их экземпляров обычно много. При этом каждая система может взаимодействовать как со множеством своих «подчинённых» систем нижнего уровня, так и со смежными системами своего уровня, а информация об объекте управления и проблемной области передаётся для агрегирования на верхний уровень.

3. Верхний уровень – обычно одна большая система для интеграции информационных потоков с нижних уровней для поддержки принятия решений в рамках всех сфер жизни.

Это положение может быть проиллюстрировано следующей диаграммой (здесь: СППР – система поддержки принятия решений, ИАИС – интеллектуализированная информационно-автоматизированная система, ИА – интеллектуальный агент).




В каждом из следующих разделов, по возможности, будут так или иначе рассмотрены все эти три уровня ИИ-систем. При этом необходимо отметить, что излагаемые в следующих разделах мысли и идеи направлены, скорее, на активизацию интуиции и творческих усилий читателя, который сможет найти возможности для применения ИИ-технологий в тех сферах жизни, которые интересуют его. Тем не менее я рекомендовал бы прочитать все разделы, так как все сферы жизни в любом случае тесно взаимосвязаны, так что внедрение ИИ-технологий в одной из них будет несомненно влиять на все другие.

## Раздел 6.1. Безопасность

Безопасность является базовой потребностью как отдельного индивидуума, так и общества в целом, поэтому государство является инструментом обеспечения безопасности внутренней и внешней. Функция обеспечения безопасности во многих случаях даже может иметь очень высокий приоритет для государства – как

в определённые периоды времени, так и в зависимости от структуры и формы государственного устройства.

Обеспечение безопасности направлено на стабилизацию общественных отношений, недопущение нарушения функционирования системообразующих элементов и институтов в самом государстве, выявление и предупреждение рисков и угроз изнутри и снаружи. Однако при выполнении функции обеспечения безопасности важно соблюдать баланс между «закрученными гайками» и возможностью для общества и государства двигаться вперёд, развиваться.

Также необходимо отметить, что у безопасности имеется множество различных аспектов – физическая, информационная, экономическая и другие, – все вместе объединяемые под термином «комплексная безопасность». Наконец, необходимо отметить, что всё то, что выше написано в отношении государства, может относиться и к более простым социотехническим системам – организациям (юридическим лицам).

Обычно при рассмотрении любого аспекта безопасности в расчёт принимается модель рисков и угроз и модель нарушителя. На основании этих моделей осуществляется стратегическое и оперативное планирование, резервирование сил и средств для предотвращения угроз, обеспечения безопасности и ликвидации последствий. Далее в соответствии с разработанными программами и планами в действие приводятся меры по обеспечению безопасности. Обычно всё вышеперечисленное основано на стандартизации и использовании зарекомендовавших себя методик и принципов. Ведь недаром говорят, что «техника безопасности выстрадана кровью и смертями».

Мы рассмотрели необходимые понятия из области безопасности, а теперь перейдём к тому, что здесь могут дать технологии искусственного интеллекта. Самое главное – это динамика в части моделей угроз и нарушителя, их постоянная адаптация под меняющиеся условия функционирования.

Ведь не секрет, что между «иммунной системой», назовём орган обеспечения безопасности так, и нарушителями постоянно ведётся «гонка вооружений». Этот биологический термин применим к системам любой природы, и не зря его использовали во время холодной войны для обозначения попыток сверхдержав нарастить потенциал своих атакующих и оборонительных систем

так, чтобы получить доминирование над противоположными системами потенциального противника.

Другими словами, гонка вооружений – это постоянные попытки со стороны нарушителя преодолеть меры безопасности, а со стороны иммунной системы предугадать и предотвратить такие попытки. Что будет, если система обеспечения безопасности основана на вырубленных в скрижалях правилах и бюрократических процедурах, которые должны неукоснительно исполняться? Правильно! Такую систему будут постоянно «хакать» разными необычными способами. Особенно в связи с развитием технологий. Так вот главной задачей искусственного интеллекта в этом случае становится создание механизмов адаптации системы безопасности к новым условиям среды, в которой она функционирует.

Не менее важными способами применения технологий искусственного интеллекта являются частные подсистемы и отдельные функции. Например, задача распознавания образов хорошо ложится на системы видеонаблюдения, добавляя в них функцию детекции инцидентов и видеоаналитики. Уже сегодня на рынке существуют готовые решения, которые позволяют методами видеоаналитики определять многочисленные классы событий – от прохождения через периметр контроля до открытого огня. Можно сказать, что детекция инцидентов – это наиболее широко применяемая технология искусственного интеллекта, развитая уже сегодня.

Но вот, к примеру, анализ больших данных и дата-майнинг в этих вопросах применяются несколько реже, хотя именно эти технологии были бы интересны в связи с развитием адаптационных механизмов, так как позволяют заранее обнаружить скрытые закономерности и использовать их для актуализации моделей безопасности.

Ещё одна технология искусственного интеллекта – использование деревьев решений для высокоуровневого принятия решений относительно методов реагирования на те или иные события, пришедшие из системы видеоаналитики и детекции инцидентов. Сейчас обычно оператор, которому система сигнализировала о наступлении события, принимает решение, что делать дальше. Но ведь это можно делать автоматически. В том числе и в рамках уже неоднократно упомянутых адаптационных механизмов.

Хорошо. Давайте рассмотрим примеры. Первый очень часто применяется на автомобильных дорогах. Представьте себе трассу длиной в две тысячи километров. Трасса идёт по безлюдным

степям. Каждые триста метров на трассе располагаются две видеокамеры высокого разрешения, которые постоянно снимают дорогу в обе стороны. Всего таких видеокамер получается более тринадцати тысяч. Если один человек может охватить взглядом 16 трансляций видеопотоков в течение двух часов наблюдения, то для непрерывного видеонаблюдения в течение полных суток требуется пятнадцать тысяч человек, работающих в шесть смен один день через три. Это слишком, не правда ли?

А с учётом того, что на трассе обычно ничего такого не происходит, всем со этим хозяйством управляются шесть операторов в смену, т. е. штат составляет чуть более пятидесяти человек, включая руководителей. А всё дело в автоматической детекции инцидентов: дорожно-транспортных происшествий, движения задним ходом, остановки, выхода на проезжую часть человека или крупного животного и даже выпадения груза и возгорания транспортного средства. И система сигнализирует оператору только в случае, если инцидент распознан и подтверждён. Ошибок второго рода практически нет, так как на одну и ту же точку автомобильной дороги в любой момент времени смотрит, как минимум, две видеокамеры, которые работают синхронно. Оператору остаётся лишь выслать на место происшествия беспилотный дрон или сразу аварийного комиссара для разбирательства на месте.

Второй пример. Представьте детскую площадку. На ней есть некоторое количество сенсорных устройств, в том числе видеокамеры с аналитикой, тревожная кнопка и датчики на опасном игровом инвентаре. Видеокамеры обозревают не только саму площадку, но и подходы к ней. А сама система не просто распознаёт инциденты и реагирует в реактивном режиме, но действует проактивно, предугадывая то, что может произойти. Например, она может зафиксировать приближение к детской площадке лица, ранее привлекавшегося к уголовной ответственности и вышедшего по условно-досрочному освобождению, у которого ещё не прошёл испытательный срок. Это может быть выяснено, например, по идентификации телефонного аппарата, который находится у такого лица. После этого система направит на телефоны людям, находящимся на площадке, тревожные сигналы о необходимости соблюдать повышенную осторожность. Тут используется анализ больших данных, в том числе завязанных на геопозиционирование.

Сегодня подобные системы уже начинают применяться в таких государствах, как Китай. Часто они завязаны на так называемый «социальный рейтинг». Да, это фактически дискриминационные системы, но, как написано выше, каждое общество само должно найти баланс между частными интересами и свободами и общественной безопасностью.

## Раздел 6.2. Полицейский искусственный интеллект

Главная задача полиции – предупреждение правонарушений и преступлений внутри общества. Полиция является как бы первым эшелоном иммунной системы общества, занимаясь борьбой с теми, кто дестабилизирует общественный порядок и устройство. Само собой разумеется, что технологии искусственного интеллекта могут дать полицейским огромные возможности. Да и что говорить – уже сегодня дают. Давайте рассмотрим несколько таких возможностей.

Во-первых, поскольку в основе деятельности полиции лежит обеспечение общественной безопасности, всё то, что было ранее описано с точки зрения интеллектуализации безопасности, относится и к полиции. В частности, предиктивная аналитика должна самым категоричным образом помочь полицейским в вопросах именно предупреждения правонарушений и преступлений. И здесь хорошая новость заключается в том, что полиция начнёт заниматься своей главной задачей, а не «бить по хвостам».

Впрочем, для этого также требуется по максимуму исключить тот самый пресловутый человеческий фактор, негативные аспекты которого постоянно вмешиваются в работу полицейских сил. ИИ-системы в этом отношении станут отличной альтернативой, так как их будет невозможно подкупить или запугать, как-то воздействовать на них, кроме как физически или информационно. Но физическая и информационная защита таких систем – это уже совершенно иной вопрос.

Итак, с точки зрения предиктивной аналитики потребуется массовое изучение больших данных, собираемых отовсюду. Здесь будут использоваться видеопотоки со стационарных видеокамер для наблюдения за оперативной обстановкой с автоматическим выявлением инцидентов. Здесь также будет использоваться интеллектуальный анализ курсирующей в сети информации

для проведения оперативно-розыскных мероприятий. Уже сегодня это делается, правда, в мониторинге ключевых слов интеллектуальности немного. Но в будущем анализ будет производиться более качественно, что как раз-таки позволит заниматься профилактикой и предупреждением преступлений.

Но предиктивная аналитика и поисковая система на основе визуальной информации может быть доступна не только в каком-либо оперативном центре, но и непосредственно на местах дислокации полицейских. Речь идёт об использовании технологии дополненной реальности совместно с ИИ-технологиями. Что, если полицейских снабдить специальными очками или системой визуализации и предоставления аудиоинформации так, чтобы они наблюдали оперативную обстановку через такое устройство? Оно могло бы сразу сверять распознанные лица окружающих с базой данных поиска либо предупреждать полицейского об опасности. И это, очевидно, серьёзно повысит уровень вовлечённости полицейского в обеспечение общественной безопасности.

Впрочем, это будет только до появления роботов-полицейских. Уже сегодня ведутся разработки и в этом направлении – роботы, которые охраняют правопорядок. Поначалу это могут быть ИИ-системы для автономного курсирования по какому-либо маршруту, видеонаблюдения и детекции инцидентов и сигнализации о них людям. Позже это уже могут быть полностью автономные системы нелетального действия, в том числе и для задержания преступников. Впрочем, до этого ещё далеко. Но наверняка будет.

А что уже есть сегодня, так это, конечно же, системы поддержки принятия решений для управления полицейскими силами и средствами в кризисных и чрезвычайных ситуациях. Сегодня пока это довольно примитивные системы в основном для предоставления оперативной и стратегической отчётности, но с накоплением больших данных они будут расти и эволюционировать в системы управления предотвращением кризисных ситуаций и реагированием на них. Такие системы будут именно управлять и принимать решения на основе самостоятельно модифицирующихся моделей. К этому всё идёт.

И в общем-то, задачей общества является тщательный контроль процесса эволюции таких систем, чтобы они не стали инструментом подавления гражданских свобод и прав, так как применение технологий искусственного интеллекта также может привести и к авторитарному методу управления обществом.

# Раздел 6.3. Военный искусственный интеллект

Ничто в области искусственного интеллекта не вызывает столько споров и не является столь противоречивой темой, как применение ИИ-технологий в военных целях. И в этом вопросе большинство исследователей и учёных занимает довольно крайнюю позицию – все исследования по возможности применения искусственного интеллекта на войне необходимо исключить или, по крайней мере, серьёзным образом ограничить и вести под контролем надгосударственных структур.

Понять эту позицию можно. Если ИИ-система – это в первую очередь автономная система, самостоятельно принимающая решения, то дать ей возможность принимать решения о жизни человека – это очень опрометчиво и, в общем-то, страшно. Особенно если учесть, что сегодня ИИ-системы не могут толком объяснить мотивы и причины своих решений. Так что если военная ИИ-система будет обладать летальным действием, то беды не миновать. И это будет самое страшное оружие массового поражения в истории человечества. Некоторые авторы угрожают, что это вообще будет последнее изобретение человечества.

Тем не менее с противоположной стороны раздаются не менее резонные доводы. Действительно, если запретить разработку военных ИИ-систем, есть высокий риск следующего развития событий – такие системы будут разрабатываться в подпольных лабораториях и государствах-изгоях. Если цивилизованные и высокоразвитые общества самостоятельно ограничат себя в развитии военных ИИ-технологий, то они заведомо поставят себя в невыгодное положение. Ведь проблема в том, что если для разработки ядерного оружия или иного оружия массового поражения предыдущих поколений необходима очень дорогая научная и технологическая инфраструктура и кадры с серьёзной подготовкой, то для разработки ИИ-систем сегодня нужен компьютер и светлая голова. И ограничить это уже невозможно – джин выпущен из бутылки, в мире опубликовано огромное количество учебной литературы и выпущено большое количество курсов по искусственному интеллекту, машинному обучению и другим технологиям.

Так что высокоразвитые общества уже просто не имеют права ограничивать себя в вопросах разработки военных ИИ-систем, так как их наличие у разных сторон будет гарантией их непри-

менения, как это происходит с ядерным оружием. Скорее всего, именно из-за этого между первыми экономиками мира уже сегодня объявлена новая гонка вооружений, ведь способов применения у военных ИИ-систем намного больше, чем у других видов оружия массового поражения, а это значит, что защита от них должна обладать соответствующим уровнем общности. Как минимум, эта новая гонка вооружений приведёт к установлению неустойчивого равновесия, на котором будет балансировать хрупкий, но всё же мир. Конечно, было бы лучше иметь устойчивое равновесие, при котором военных ИИ-систем не было бы в этом мире, но пасту в тюбик обратно уже не засунуть.

Размышления на эту тему приводят к альтернативной идее, что развитые государства должны сосредоточиться на разработке военных ИИ-систем нелетального действия. Действительно, если на поле боя столкнутся боевой искусственный интеллект и живая сила и техника другой стороны, то первая сторона должна своей мощью подавить вторую, не нанося урона живым людям. Воевать должны технологии, а не люди, и если в каких-то государствах до сих пор считают иначе, то их необходимо принуждать к миру именно такими нелетальными боевыми ИИ-системами. Хотя, скорее всего, это вопрос ещё довольно отдалённого будущего.

Сегодня же вариантов применения искусственного интеллекта в военном деле довольно много, и их имеет смысл перечислить.

1. Обеспечивающие системы на поле боя, например роботы, которые эвакуируют раненых или выполняют курьерские функции.
2. Видеоаналитика, совмещённая с системами дополненной реальности для использования непосредственно в бою.
3. Рои автоматизированных или автономных дронов, участвующих в боевых действиях.
4. Системы поддержки принятия решений на тактическом, оперативном и стратегическом уровнях.

В силу особого положения этой темы и её высокой спорности далее в этой книге она рассматриваться не будет.

## Раздел 6.4. Транспорт

Транспорт – это одна из базовых систем любой экономики. Без транспорта и транспортной доступности было бы очень сложно вести экономическую деятельность, особенно на больших

масштабах расстояний. Поэтому транспортные системы шести модальностей всегда находятся в зоне повышенного внимания практически любого государства.

Шесть модальностей, про которые было упомянуто выше, – это автомобильный, железнодорожный, речной, морской, воздушный и трубопроводный виды транспорта. В развитых экономических системах все они очень востребованы и широко используются, в том числе в рамках мультимодальных перевозок и в рамках интеграции между государствами, макрорегионами и даже континентами.

Где же технологии искусственного интеллекта могут быть применены, и как они могут улучшить показатели транспортной отрасли? Начнём, как это у нас водится, с самого верхнего уровня. Можно ответственно заявить, что как раз в области транспорта с использованием искусственного интеллекта более или менее всё нормально, так как практически все акторы транспортной отрасли так или иначе занимаются этим в рамках развития так называемых интеллектуальных транспортных систем.

Работы в этом направлении ведутся, быть может, с конца 1990-х годов, и к настоящему моменту имеется огромное количество как национальных стандартов, так и различных конкретизированных технологий, которые используют отдельные элементы и технологии искусственного интеллекта.

Дадим определение интеллектуальной транспортной системе. Есть много различных вариантов, но я придерживаюсь своего собственного. Итак, интеллектуальная транспортная система – это интегрированная автоматизированная система, которая, используя интеллектуальные и инновационные методы организации и управления, предоставляет участникам транспортной отрасли сервисы по координированию, планированию, информированию, повышению уровней безопасности и эффективности использования транспортных сетей. Традиционно интеллектуальные транспортные системы относятся только к автомобильному транспорту, однако разносторонняя интеграция с транспортом иных модальностей подразумевается.

Теперь рассмотрим, что же делает интеллектуальную транспортную систему именно интеллектуальной. Ведь её функции вполне могут исполняться обычными автоматизированными

системами управления. Поэтому функциональность такой транспортной системы должна обладать рядом свойств.

Во-первых, ИТС должна уметь интерпретировать получаемые данные «на лету». Это как раз и есть тот самый анализ больших данных, поиск и выявление закономерностей в состояниях объекта управления и использование этих находок в работе.

Во-вторых, ИТС должна уметь диагностировать своё собственное состояние. Самодиагностика – это важный элемент интеллектуальности, и он, в свою очередь, также основан на постоянном анализе и интерпретации данных о состоянии своих элементов.

В-третьих, ИТС, конечно же, должна осуществлять мониторинг текущего состояния объекта управления.

А в-четвёртых, она должна уметь моделировать и прогнозировать будущие состояния объекта управления и свои собственные.

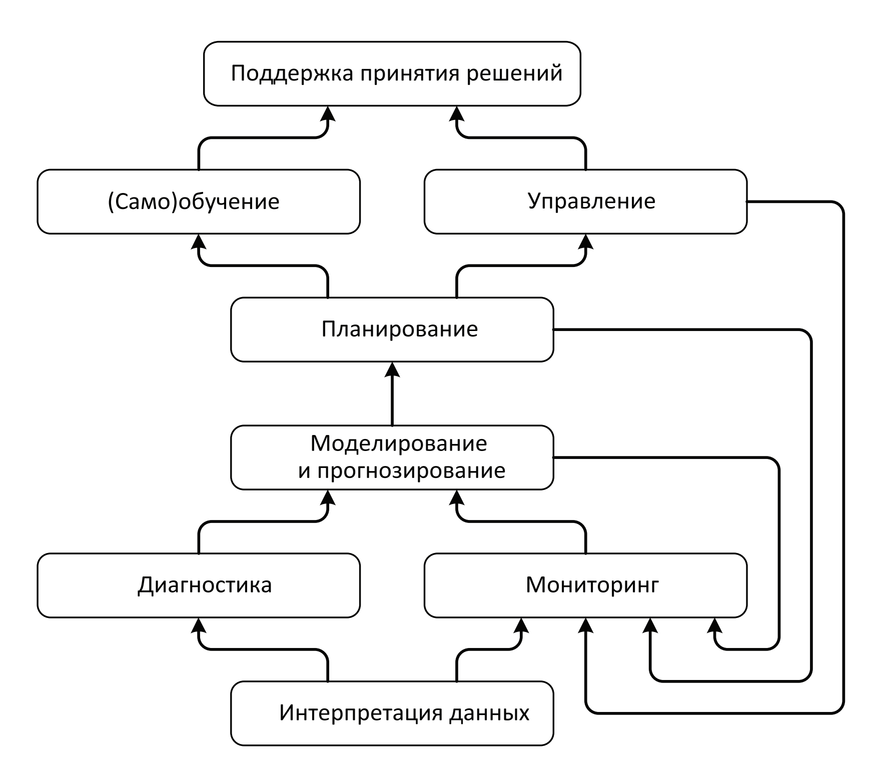
В-пятых, ИТС должна уметь планировать свои реакции на будущие состояния. Планирование используется в функциях ИТС для подготовки будущих действий на основании прогнозов с дальнейшим сравнением прогноза, плана и факта с передачей отклонений по обратной связи в модель, используемую для прогнозирования, для её актуализации и перекалибровки. Тем самым готовится почва для обучения системы. Кроме того, планирование является важнейшей предпосылкой управления.

Соответственно, в-шестых, ИТС должна уметь обучаться, что может происходить как автоматически, так и на основании наблюдения за действиями персонала (дедуктивно).

В-седьмых, ИТС должна уметь управлять своим объектом в разных режимах, переключаясь в них на основании своего прогноза и получаемых фактических данных. Очень важно, чтобы в конечном итоге ИТС использовалась для управления в кризисных ситуациях, что позволит смягчить последствия.

И наконец, в-восьмых, ИТС должна быть автономной. Пока степень автономности низкая, ИТС используется для поддержки принятия решений. Но чем больше становится интеллектуальности, тем более автономной должна быть ИТС. И такой аспект функциональности ИТС должен стать той целью, к которой необходимо стремиться при проектировании и реализации системы.

Таким образом, все перечисленные интеллектуальные аспекты функциональности складываются в следующую диаграмму.




Итак, максимально возможное применение методов искусственного интеллекта для реализации перечисленных аспектов интеллектуальности, особенно в части обработки больших данных, обучения и поддержки принятия решений, будет залогом того, что транспортная система действительно станет интеллектуальной.

Реализация подобной системы, конечно же, требует огромных затрат и большого объёма времени. Здесь может быть два подхода – сверху и снизу. Но, как обычно, превалирует подход снизу, поскольку уже существует значительное количество «низовых» систем, охватывающих отдельные функции и аспекты транспортной отрасли. Давайте к ним и перейдём.

Например, рассмотрим уже полюбившиеся нам автономные автомобили. А что, если система управления маршрутами для них будет централизованной, и это получится как бы «царица» с роем муравьёв? В этом случае роль «царицы» должна выполнять интел-

лектуальная система планирования и контроля, которая находит решение огромного количества задач движения по графу дорог с учётом временных ограничений. Если такое реализовать, то можно будет забыть о пробках, светофоры будут не нужны, движение будет безостановочным, количество автомобилей сократится, количество выбросов в атмосферу сократится ещё больше. В общем, будет очень хорошо.

Но такая система – дело довольно далёкого будущего. Сегодня же, когда автономных автомобилей на дорогах общего пользования ещё нет (или же они только-только появляются), надо как-то оптимизировать движение как на отдельных перекрёстках, так и на сети автомобильных дорог. Это достигается различными методами, самыми интересными и новыми из которых являются методы адаптивного управления.

Адаптивное управление – по сути, это подстройка времени переключения фаз светофоров в соответствии с потребностями на пересекающихся направлениях с учётом пешеходного трафика. Для одного перекрёстка решить эту задачу довольно непросто, но вполне возможно. Для зоны адаптивного управления – очень тяжело, но тоже возможно. Чем больше зона и чем больше в ней светофорных объектов, тем тяжелее. Ну и на полной улично-дорожной сети таких городов, как Москва, решить эту задачу невозможно практически, а про аналитическое решение никто не знает.

Итак, главной задачей искусственного интеллекта в области транспорта будет полное высвобождение человека из управления и организации движения на низовом уровне. Можно сказать, что будущая транспортная система будет решать оптимизационную задачу прокладывания маршрута из заданной точки А в точку Б с минимизацией времени движения. Заказчик услуг мобильности будет только указывать начальную и конечную точки маршрута и возможные ограничения, а система сама доставит пассажиров или груз точно, безопасно и оптимально.

И искусственная интеллектуальная система, которая будет решать эту задачу, будет основывать свои действия на постоянном обучении и адаптации своего поведения на основе больших транспортных данных, собираемых отовсюду: с борта транспортных средств, с элементов дорожной инфраструктуры, с различных устройств пешеходов и пассажиров, из результатов непрерывного моделирования.

# Раздел 6.5. Энергетика

В области энергетики также есть различные широкие возможности применения ИИ-технологий. Начнём с очень частных и конкретных задач и постепенно дойдём до глобальных.

В последнее время искусственные нейронные сети, собранные в генеративно-состязательную схему, используются для автоматического проектирования разных устройств и агрегатов. Для этого такой связке нейронных сетей определяют набор требований и ограничений, после чего начинается процесс поиска и подбора. Довольно быстро они находят интересные решения, полностью подходящие под заданные условия и выполняющие требуемые функции. Эта модель используется для генерации изображений, написания текстов, создания рецептов, рисования выкроек и, кстати, проектирования узлов и приборов. Последнее вполне можно использовать и для электроэнергетики.

Например, есть опоры для линий электропередач, которые вполне можно спроектировать таким особым методом, поставив в качестве условий стоимость и вес материалов, парусность, различные нагрузки и т. д. Не факт, что в результате получится такая же опора, что обычно проектируется человеком, – есть примеры проектов узлов, спроектированных таким методом, при взгляде на которые кажется, что их создал «внеземной разум». В какой-то мере так оно и есть. Но они полностью выполняют свои функции, а при усилении требований и ограничений ИИ-система может найти решение намного быстрее человека, который вообще может опустить руки.

Эту технологию автоматического проектирования можно использовать не только для опор линий электропередач, но и вообще для всего. Например, можно проектировать узлы и агрегаты, входящие в системы и комплексы на электростанциях. Можно проектировать провода, переключатели, предохранители и всё-всё-всё. В общем, есть довольно широкий простор для применения генеративно-состязательной архитектуры искусственных нейронных сетей.

Идём дальше. Есть задача проектирования и создания «умных» сетей электроснабжения. Что в этом случае понимается под словом «умный»? Как минимум, три аспекта.

Во-первых, при проектировании и прокладывании сетей необходимо решать задачу оптимизации маршрутов с учётом огромного количества факторов. Многокритериальная оптимизация

в пространствах высокой размерности – это то, что искусственный интеллект умеет делать очень хорошо. Так что методы поддержки проектирования при помощи ИИ-технологий смогут или самостоятельно, или же в паре с проектировщиками и конструкторами создавать оптимальные трассировки линий электропередач.

Во-вторых, когда у нас уже есть построенная сеть электроснабжения, всегда возникает задача перебалансировки нагрузки, причём в динамическом режиме. Дело в том, что потребители никогда не сидят на одном месте – высокозатратные с точки зрения потребления энергии производства закрываются и открываются, большие массы людей переезжают с места на место, в зависимости от времени суток, дня недели, наличия локальных праздников и сезона года меняется потребность в электроэнергии в различных локациях, куда доводится электричество по сети. Несмотря на то что эта задача в целом решена, скорее всего, при помощи методов искусственного интеллекта можно ещё больше оптимизировать этот процесс так, чтобы терялось ещё меньше энергии. А это значит, в свою очередь, что нагрузка на экологию будет меньше. Это здорово.

В-третьих, мы вообще сможем начать планировать потребности в электроснабжении на локальном и глобальном уровнях, в результате мы сможем более оптимально и эффективно инвестировать в отрасль. Кто может сейчас ответить, сколько мощностей генерации где и каких необходимо заменить на альтернативные источники энергии в пятилетней перспективе, чтобы максимально эффективно оптимизировать отрасль? Никто. Искусственный интеллект при наличии всех необходимых знаний и собранных данных смог бы.

Все эти задачи вполне решаются глобальной системой поддержки принятия решений, которая должна управлять всей отраслью генерации и распределения электроэнергии. Такая система должна собирать низкоуровневую информацию со всех предприятий, задействованных в отрасли, – от всех электростанций всех типов, от всех подстанций распределения, от всех сетей. И она должна заниматься стратегическим управлением, имея в наличии модель объекта управления, планы по его модернизации и перспективы изменения той среды, в которой работает вся система электроснабжения. Когда (и если) такая система поддержки принятия решений будет построена, мы постепенно сможем отдавать ей все вопросы в самостоятельное решение, и она станет автоматической системой принятия решений. И тогда всё будет хорошо.

## Раздел 6.6. Сельское хозяйство

Как обычно, начнём с базовых отраслей экономики. Продукты питания создаются в рамках различных сегментов сельского хозяйства. И сельское хозяйство – это та отрасль, где технологии искусственного интеллекта уже сегодня находят своё применение. Например, если беспилотные автомобили только-только начинают выходить на автомобильные дороги общего пользования, то в сельском хозяйстве беспилотные комбайны и другие сельскохозяйственные машины используются уже сегодня. Такие машины не просто ездят по полям и выполняют свои функции, но и осуществляют постоянный мониторинг своего состояния, синхронизируются с системой технического обслуживания и ремонта, заранее заказывают изнашивающиеся детали и агрегаты, чтобы обеспечить непрерывность процесса своей работы.

Но становятся беспилотными не только комбайны и трактора. Вообще, появляются новые сельскохозяйственные роботы, которые берут на себя механический труд человека. Например, робот, который выпалывает сорняки: он работает двадцать четыре на семь, не спит, не уходит в отпуск, а выпалывает сорняки так, что гербициды не требуются. Это в том числе повышает «экологичность» продуктов. Или вот робот-пастух, который знает всех своих подопечных в лицо и наперечёт. Но при этом нельзя сказать, что это такое антропоморфное существо с кнутом в железных руках. Это распределённая ИИ-система с компьютерным зрением и другими технологиями. И вообще, в основном внутри таких сельскохозяйственных роботов используются технологии восходящей парадигмы.

Но сельское хозяйство – это серьёзная штука, которая в современных условиях довольно сильно зарегулирована. Практически любой аспект деятельности подчиняется различным нормативно-правовым актам, и фермер зачастую вынужден заниматься только различными бумагами, а не выращивать растения или животных. И часто такие нормативные требования к деятельности фермеров и реализуемой ими продукции противоречивы, запутанны и неадекватны изменяющимся условиям современной экономики. Так что бедному фермеру очень сложно разобраться во всех этих тонкостях, а за его спиной уже маячат различные контрольно-надзорные органы и почему-то в подавляющем большинстве случаев с кнутом. Меня удивляет, что в таких условиях вообще кто-то занимается сельским хозяйством.

В этом вопросе на помощь придёт система поддержки принятия решений в области сельскохозяйственной юриспруденции, которая должна взять на себя все вопросы оформления лицензий, разрешений, справок и т. д. Будучи основанной на нисходящей парадигме и работающей с юридическими и сельскохозяйственными знаниями, она снимет всю рутину по работе с бюрократией. Насколько мне известно, примеров таких систем поддержки принятия решений в мире ещё нет.

## Раздел 6.7. Лёгкая промышленность

При рассмотрении возможностей искусственного интеллекта в лёгкой промышленности (во всех её аспектах) начнём с того, что ИИ-системе вполне можно поручить создавать новые фасоны и модели одежды, в том числе и модных брендов. Вот, к примеру, идея для стартапа – открыть дом моды и раструбить на весь мир, что модели в нём создаются искусственным интеллектом. Впрочем, почему «раструбить»? Ведь это вполне легко можно реализовать, причём именно на гибридной схеме. Если скормить нейронной сети глубинного обучения большое количество выкроек для моделей одежды от известных и не очень кутюрье, а потом сделать генеративно-состязательную схему, то в итоге получится, как кажется, довольно интересно. А вот доводить до ума выкройки, созданные этой парочкой нейронных сетей, должна символьная ИИ-система, в базе знаний которой содержатся эвристические правила мира моды и в том числе правила масштабирования выкроек для разных размеров.

Такая система могла бы генерировать большое количество модных моделей хоть каждый день, но ей можно было бы поручить разрабатывать, к примеру, сезонные модели для разных стран и регионов мира. Думаю, что при грамотной маркетинговой политике и продвижении успех у такого стартапа был бы большой. Как минимум можно будет продаться какому-нибудь из больших домов моды. Насколько мне известно, подобных систем ещё нет ни у кого, так как считается, что творчество – это прерогатива человека, хотя искусственный интеллект уже давно научился создавать нормальные креативные штуки. Всё дело в генеративно-состязательной архитектуре.

Идём дальше. Ну ладно, вот умненькая ИИ-система, подвизавшаяся модисткой, создала множество моделей одежды, каждую

из которых так и хочется примерить. Но их ещё нигде нет, в примерочную в модном бутике не зайти. А некоторым мужчинам, к примеру, и вовсе не нравится ходить по модным бутикам, но выглядеть модно тоже хочется. Так почему бы не предоставить таким людям виртуальные примерочные. Несмотря на то что это, кажется, больше относится к технологиям виртуальной и дополненной реальности, без искусственного интеллекта здесь не обойтись.

Примерять на манекен неинтересно даже в виртуальной примерочной. Вернее даже, особенно неинтересно в виртуальной примерочной использовать манекен. Что мешает нам осуществить трекинг и трёхмерную съёмку тела конкретного клиента, построить 3D-модель и использовать её в виртуальной примерочной? Да легко. Уже сегодня есть такие технологии для построения подобного рода моделей и трекинга скелета, чтобы можно было заставить виртуальную модель принимать практически любую позу. А те же самые генеративно-состязательные нейронные сети смогут одеть такую 3D-модель клиента в выбранную модель одежды. Сделать это несложно. Впрочем, сегодня это будет достаточно долго даже на серьёзных вычислительных мощностях в каком-нибудь центре обработки данных, не говоря уже о персональном компьютере клиента. Но технологии развиваются, и вскоре нас ждёт вал подобных предложений.

Но если всё это возможно и вроде бы так просто, то почему бы не вспомнить о персонализации? Ведь здесь постоянно пишется о том, что искусственный интеллект – это про персонализированные сервисы, услуги которых можно будет оказывать каждому конкретному клиенту на основе его «больших данных». Персонализированный маркетинг, персонализированная медицина, персонализированное образование – это то, что нас ждёт в ближайшем будущем. Почему бы не приплюсовать к этому набору и персонализированный пошив одежды? Можно…

Персонализированный пошив одежды есть и сегодня – это дорого и долго, зато очень удобно и ценно для важных клиентов. Модные кутюрье, обслуживающие таких важных клиентов и шьющие им одежду в персональном режиме, зарабатывают хорошие деньги. Но боюсь их огорчить – скоро их место займут роботы, которые будут работать за электричество.

Итак, вполне можно устроить персонализированный пошив одежды, причём опять же клиент делает трекинг своего тела, далее строится 3D-модель, и ИИ-система снимает все важные

показатели для выкройки. Затем у клиента берётся интервью на тему того, что бы он хотел получить – то ли деловой костюм с иголочки, то ли костюм на маскарад. В результате формируется техническое задание, которое и исполняется ИИ-системой в виде конкретных персонализированных выкроек для этого конкретного клиента. Вуа-ля!

Но если у нас есть такие технологии, то почему бы нам не использовать для пошива одежды всякие «умные» штуки? При помощи квантовых технологий можно будет создать умные материалы и всякого рода сенсоры, которые будут осуществлять непрерывный и неинвазивный мониторинг жизненных показателей того, кто носит такую одежду. Ведь это получится сбор больших данных, которые затем могут использоваться и для медицинских целей, и для моды, и ещё много где.

Однако давайте теперь перейдём от небольших задач искусственного интеллекта в мире моды к глобальным задачам лёгкой промышленности. Существует целая отрасль, в которой есть предприятия и производственно-экономические цепочки, составленные из них. И отраслью в целом, и отдельными предприятиями могут стратегически управлять системы поддержки принятия решений. Такие системы могут собирать большие данные, прогнозировать различные ключевые показатели и планировать работу подведомственных предприятий так, чтобы эффективно использовать ресурсы, не тратить лишние материалы и обеспечивать каждого персонализированным сервисом точно в срок. Именно к этому идёт экономика во всех сегментах и отраслях, почему бы ей не прийти к этому и в лёгкой промышленности?

## Раздел 6.8. Персонификация диетологии

В рамках диетологии и вообще технологии пищевого производства есть широчайшие возможности по применению самых разных методов искусственного интеллекта всех парадигм. Давайте кратко рассмотрим наиболее интересные, с моей точки зрения, варианты применения ИИ-технологий в этой области.

Например, меня всегда интересовало, кто придумывает рецепты различных блюд. Раньше в стародавние времена, по всей видимости, это происходило примерно так же, как создавались народные сказки и подобные образцы народного творчества, – из поколе-

ния в поколение передавались мемы. Потом, когда потребовалось массовое производство продуктов питания, появились технологи пищевого производства, которые и начали придумывать рецепты. А сегодня уже появляются нейросетевые ИИ-решения, которые генерируют рецепты. Это очень интересная задача, которая решается несколькими коллективами, и иногда найденные рецепты довольно интересны. Впрочем, я сам не пробовал.

Вместе с тем само по себе пищевое производство – это сегодня большое предприятие с развитыми технологическими и рабочими процессами. Само собой разумеется, что подобным предприятием можно управлять при помощи системы поддержки принятия решений, которая собирает большие данные со всех подчинённых процессов и занимается прогнозированием и стратегическим планированием бизнеса. Опять же, эта задача не решена практически нигде, хотя символьные и логические ИИ-системы в будущем станут автономными топ-менеджерами на любом предприятии.

Представим: мы придумали рецепты и сделали продукты питания, а что дальше? Их надо приготовить и подать клиентам, например, в ресторане. С этим тоже могут справиться роботы, и у нас будут роботы-повара и роботы-официанты. В этом направлении тоже наблюдаются подвижки, и сегодня есть полностью автоматические рестораны, куда человек заходит откушать, делает заказ, и его выполняет и подаёт полностью роботизированная система. Фактически это будет то будущее массового общественного питания, к которому идёт мир.

Что ж, искусственный интеллект – это про персонализацию сервисов. Технологии искусственного интеллекта, позволяя обрабатывать самую персональную информацию о клиенте, предоставляют наиболее подходящие конкретно этому клиенту услуги. Не обошла персонализация и питание. Ведь если рассмотреть персонального помощника, который следит за тем, что употребляет его «хозяин» в качестве продуктов питания и питья, то он сможет не только подсчитывать калории и другие характеристики питания, но и давать персонализированные рекомендации, что и когда есть, чтобы быть здоровым и иметь высокое качество жизни. К этому всё идёт, такие персональные помощники разрабатываются и уже используются, правда, в несколько усечённом виде.

В будущем подобные персональные помощники в области диетологии будут готовить для человека конкретное меню на каждый день и отслеживать, чтобы человек питался правильно. Это

будет переход именно к персонализированной диетологии, когда все факторы жизни будут учитываться при планировании питания. А на глобальном уровне это приведёт к серьёзной экономии ресурсов, так как каждый будет питаться так, как требуется, не больше и не меньше. По цепочке это повлечёт снижение нагрузки на экологию и увеличение уровня и качества жизни.

## Раздел 6.9. Медицина

Медицина как одна из областей доверия очень консервативна и жёстко зарегулирована. В этой области вряд ли в ближайшее время примут к использованию автономные искусственные интеллектуальные системы без возможности объяснения своего поведения и принятых решений. Это касается любой области – диагностики, выбора и планирования лечения, прогнозирования развития состояний, проектирования новых лекарственных средств и т. д. Всё это требует тщательного контроля, а потому в эксплуатацию будут приниматься только XAI-системы, и то в первое время под тщательным контролем человека.

Вместе с тем сегодня с развитием генных технологий и других биотехнологий медицина также переживает небывалый подъём и прорыв в новых областях. Во многих развитых странах набирает силу тренд на переход к точной (precision) и персональной (personal) медицине, суть которой составляет так называемый OMICS-профиль пациента, состоящий из данных геномики, протеомики, метаболомики и нескольких других срезов биохимических процессов в организме. Омик-данные, собранные для конкретного пациента, используются в качестве основы для принятия решений по диагностике и терапии. В частности, сегодня на многих лекарственных средствах используется фармакогеномная маркировка, которая помогает уточнить, насколько конкретный препарат подходит пациенту с генетической точки зрения. Обычно указываются гены цитохрома P450, но встречаются и другие. И теперь любой человек, который владеет своей генетической информацией, может осознанно принимать промаркированные лекарства, так как при наличии описанных или даже неописанных мутаций в определённых генах решение о приёме препаратов, взаимодействующих с белками этих генов, принимается более осознанно.

Но даже если и не рассматривать сложные биохимические данные, которые выявляются у пациента в результате специализированных анализов, например секвенирования полного экзома, то принципы персональной медицины всё равно можно применять. Например, в инструкции каждого лекарственного средства есть разделы «Противопоказания» и «Побочные эффекты», и если на первый из указанных разделов ещё обращают внимание, особенно в сложных случаях, то на второй практически никогда. Но в нём часто указаны довольно серьёзные побочные эффекты, и в редких случаях даже летальный исход от применения лекарства (хотя это и очень редко). Отмечается ряд очень часто встречающихся побочных эффектов типа аллергии, сыпи, тошноты или головокружения. Если взять более или менее серьёзный препарат, действующий на системном уровне, то эти четыре побочные реакции наверняка будут присутствовать в его описании.

Для примера можно рассмотреть типового пациента, который начал принимать выписанный врачом препарат и через пару недель почувствовал недомогание, которое проявлялось в слабости, подташнивании и головокружении при резком вставании. Обращение к инструкции показало, что эти побочные эффекты действительно перечислены в ней, причём некоторые даже с пометкой «редкие». Но при этом у пациента в анамнезе уже были эпизоды головокружения и ортостатической гипертензии, сопровождавшейся тошнотой. И уже даже не важно, с чем это может быть связано – с генетическими особенностями, особенностями обмена веществ или ещё чем-либо, – данные анамнеза подсказывают, что пациент предрасположен к проявлению подобных побочных эффектов препаратов, поэтому при назначении врач мог бы принять это во внимание и выписать аналогичный по прямому терапевтическому эффекту препарат, но имеющий иной принцип действия. Однако, к сожалению, подобное практически никогда не происходит. Особенно принимая во внимание то, что регламент приёма пациентов у врачей становится с каждым годом всё жёстче и жёстче в части временных рамок.

При этом надо отметить, что сбор медицинских данных пациента становится легче год от года. Люди используют носимые устройства для сбора биометрических данных, которые синхронизируются с медицинскими приложениями и собираются в облачных хранилищах. Простые смартфоны могут фиксировать огромное количество информации об активности человека и некоторых его

жизненных показателях. Использование специальных медицинских приложений типа дневника здоровья либо банальных чат-ботов для фиксации различных событий, касающихся здоровья, – это становится трендом для людей, ведущих здоровый образ жизни.

Итак, большие медицинские данные собираются постоянно и повсеместно. При этом методы анализа больших данных тоже получили серьёзное развитие вместе с методами искусственного интеллекта, так как интеллектуальный анализ данных – это одна из типовых задач машинного обучения в частности и искусственного интеллекта в целом. Современные достижения в области интеллектуального анализа данных основаны не только на статистических методах, как это было ещё два десятилетия назад, но и на использовании искусственных нейронных сетей, генетических алгоритмов и других новых методик.

Это значит, что методы искусственного интеллекта могут самым серьёзным образом помочь максимально распространить концепцию персональной медицины. Здесь важным фактором является не только то, что искусственная интеллектуальная система будет тщательно мониторить все возможные показатели пациента, но и то, что она работает 24/7 без выходных и отпуска, к ней не надо записываться, у неё не бывает плохого настроения и прочих эмоциональных реакций, так свойственных человеку. Поэтому медицинские искусственные интеллектуальные системы, по всей видимости, могут стать очень серьёзным трендом в ближайшие годы, причём степень их доступности для широких масс населения будет достаточно высокой, так как стоимость разработки и обслуживания таких систем во время эксплуатации со временем будет только снижаться.

Так появляется необходимость использования концепции объяснимого искусственного интеллекта, так как здравоохранение, вне всяких сомнений, является сферой доверия. Системы, основанные на знаниях, с гибридной архитектурой, станут развитием экспертных систем полувековой давности. Уже сегодня такие системы под названиями «медицинская информационная система» и «медицинская система поддержки принятия решений» начинают повсеместно использовать в медицинских учреждениях (и в первую очередь в частных) и целых системах здравоохранения на уровне государств.

Гибридная архитектура таких искусственных интеллектуальных систем состоит из двух уровней. На первом уровне нахо-

дятся модули, основанные в том числе и на методах восходящей парадигмы для интеллектуального анализа данных. Как было указано ранее, именно искусственные нейронные сети являются очень эффективным методом для дата-майнинга. Но их главная проблема – невозможность объяснить полученные результаты. Однако гибридным системам не требуется объяснения результатов, полученных на нижнем уровне анализа. Эти результаты в уже обработанном для символьной математики и логики виде подаются на второй уровень, где работает интеллектуальная машина вывода, уже готовящая решения для врача на основе правил в базе знаний. И эта машина вывода вполне может объяснить, как было получено то или иное решение. В этом главный смысл и отличительная особенность гибридного подхода: на нижнем уровне анализируются данные и получаются необходимые для вывода результаты, а на верхнем уровне эти результаты используются в выводе для принятия решений, которые могут быть объяснены.

Здравоохранение – это сфера жизни высшего уровня, и в любой развитой экономической системе она должна быть. Но если образование и наука направлены в будущее и на развитие общества, то предназначение системы здравоохранения заключается в стабилизации настоящего. Несмотря на то что биологические системы имеют высокую степень адаптации, иногда они подвержены «болезням», т. е. нарушению гомеостаза и переходу в патологический режим функционирования, что может привести к необратимым пагубным последствиям. Здравоохранение необходимо, чтобы каждый член общества и всё оно в целом были в наименьшей степени подвержены болезням.

Современная система здравоохранения довольно развита. Можно упомянуть несколько примеров того, как человечество полностью победило некоторые болезни. Вирус чёрной оспы существует только в двух серьёзно охраняемых лабораториях, и уже подумывают, чтобы полностью уничтожить его. Современным детям уже не делают прививки от оспы. Лекарства проектируются в зависимости от функции, которые они должны оказывать на белки и другие вещества внутри организма. Тем не менее есть серьёзные недостатки. Самый главный – система не успевает за современными технологиями. Последние достижения в области генетики, эпигеномики, протеомики, анализа больших данных и даже 3D-печати уже сегодня позволяют перейти на полностью персонифицированный подход к каждому конкретному

пациенту. Особенно в тяжёлых случаях, когда некоторые пациенты выбиваются из статистических шести сигм. А современные правила работы врачей как раз основаны только на статистическом анализе. Речь идёт лишь об официальной медицине, здесь не рассматриваются какие-либо альтернативные методы.

В следующих областях здравоохранения применимы как методы искусственного интеллекта, так и концепция персональной медицины:

- фармакопея;
- диагностика;
- мониторинг состояния пациента;
- планирование терапии;
- прогнозирование развития состояния пациента;
- развитие системы здравоохранения на основе собранных больших данных.

Из перечисленного только последний пункт как таковой не входит в состав персональной медицины, однако он важен для её развития, так как сбор данных, их обработка и анализ, рациональные выводы и изменения в системе здравоохранения непосредственно влияют на здоровье и качество жизни каждого, кто пользуется медицинскими услугами (и косвенно даже на тех, кто не пользуется).

Фармакопея – это наука и искусство приготовления лекарственных форм для лечения пациентов. В давние времена фармацевт готовил лекарство непосредственно для больного, и можно сказать, что в те непросвещённые времена эта область была ближе к персонализации, чем в сегодняшний просвещённый век стандартизации и унификации. Тем не менее если изучить OMICS-профиль пациента, то на основе анализа его биохимических данных можно приготовить конкретизированное лекарство, которое будет точечно воздействовать на причину патологического состояния. Сегодня всё больше и больше информации появляется о метаболических путях, учёные и медики разбираются всё глубже в том, как сложные биомолекулы трансформируются в организме, какие каналы в клеточных мембранах имеются у разных типов клеток и какие рецепторы есть на них, как осуществляется химическая сигнализация между клетками и т. д. – каждый год на эту тему печатаются десятки и сотни тысяч статей. При этом развиваются и методы аддитивных технологий в медицине и фармакопее, так что в скором будущем печать персонально

спроектированных лечебных молекул и препаратов на их основе будет осуществляться на специализированных 3D-принтерах.

Диагностика – это выявление как можно более глубоких причин патологического состояния организма на основе наблюдаемых симптомов и результатов анализов. Чаще всего причины патологии лежат в области биохимии, поэтому именно персональная медицина и OMICS-профиль каждого конкретного человека помогут лечить болезни, а не подавлять симптомы, как это обычно делается в современной медицине для большинства болезней. И как полагается, методы искусственного интеллекта помогут более тщательно производить диагностику и учитывать при этом все возможные релевантные данные, собираемые с пациента носимыми устройствами и вводимые пациентом или наблюдающими за ним людьми. В этой области применимы как методы восходящего подхода, так и нисходящие методы искусственного интеллекта.

Мониторинг состояния пациента – это процесс отслеживания состояния человека и его жизненных показателей и важных для диагностики параметров внутренней среды организма. Для более эффективного лечения необходимо обеспечить как можно более частый мониторинг – в идеале непрерывный (именно поэтому в отделениях реанимации и интенсивной терапии пациенты лежат подключёнными к системам непрерывного мониторинга). Раньше и пока ещё сегодня для большинства людей мониторинг осуществляется вручную врачом или самим пациентом при периодическом осмотре, однако всё чаще, как уже упоминалось, для этих целей используются носимые устройства и другие гаджеты. И концепция персональной медицины чуть ли не требует именно частого или непрерывного мониторинга для постоянной адаптации плана лечения к изменяющейся ситуации.

Планирование терапии осуществляется на основе результатов предыдущего процесса мониторинга. Конечно, в теории этот процесс должен быть постоянно адаптирующимся к изменяющимся условиям внутренней среды организма человека. Но на практике планирование лечения осуществляется только при редких визитах к врачу, да и то на основании стандартных протоколов лечения конкретного заболевания (часто набора симптомов, а не причин), принятых в системе здравоохранения. Так что методы искусственного интеллекта и в этом вопросе приведут к возникновению персонализированного сервиса.

Прогнозирование развития состояния пациента. Само собой разумеется, что для повышения степени адаптивности программы лечения необходимо не только учитывать постоянно меняющиеся условия и параметры внутренней среды организма человека, но и делать прогнозы относительно развития его состояния. Соответственно, искусственная интеллектуальная система должна иметь внутри себя несколько моделей – пациента, заболевания, набора лекарственных средств и системы здравоохранения, чтобы делать прогнозы на основе этих моделей. А сам процесс должен быть устроен таким образом, чтобы план и прогноз всегда сравнивались с наступившим фактом, что используется для адаптации и обучения самой искусственной интеллектуальной системы.

Наконец, развитие системы здравоохранения на основе собранных больших данных – это, как сказано выше, важная составляющая среди всех процессов, так как позволяет адаптировать саму систему здравоохранения к изменяющимся условиям и постоянно улучшать те сервисы, которые оказывает персональная медицина. Ведь на основе сбора больших данных как от всех пациентов, так и от искусственных интеллектуальных систем, которые предоставляют пациентам сервисы, происходит объединение баз знаний, и вся система здравоохранения становится как бы «умнее» с каждым принятым решением каждой индивидуальной медицинской искусственной интеллектуальной системой.

В прессе и академических изданиях неоднократно упоминались уже известные системы искусственного интеллекта, про которые время от времени говорят в новостях в том аспекте, что они, дескать, лучше диагностируют какие-нибудь специфические патологии, сигналы о которых видны на рентгеновских снимках, или на магнитно-резонансной томографии, или на кривых электрокардиограммы или электроэнцефалограммы. В общем, если существуют некоторые специфические данные – фотоснимок или какая-то временная последовательность значений, – то искусственная интеллектуальная система с нейронными сетями сможет найти в них либо скрытые закономерности, либо определить наличие проблем, если нейронная сеть была обучена именно на это. На этом основаны решения, которые изучают снимки на предмет наличия на них, например маркёров опухолей.

Следующая важная задача – изучение непрерывно передаваемых с носимых устройств данных на предмет наличия в них предикторов каких-либо кризисных ситуаций. Это очень важно

для лиц с повышенным риском развития инфаркта или для пациентов с эпилепсией, в особенности с ночными приступами и приступами во сне. И здесь носимое устройство, которое постоянно замеряет некоторый набор жизненных показателей, отправляет их в искусственную интеллектуальную систему мониторинга, в которой осуществляется прогнозирование трендов и рождается сигнал, что через некоторое время может произойти кризис. В описанных случаях, в принципе, объяснение не особо и нужно. В первом случае результат проверяется и перепроверяется врачом, поэтому ему достаточно только просигнализировать о том, что на снимке, возможно, имеются какие-то проблемы. А во втором случае ложный положительный сигнал не будет иметь плохих последствий, так как в таких случаях лучше уделить больше внимания и лишний раз не пропустить ложное срабатывание, чем недосмотреть (впрочем, остаётся чисто психологическая проблема человека: снижение внимания из-за частых ложных срабатываний). Поэтому для устройств такого типа сегодня идёт борьба за снижение количества и частоты ошибок второго рода, т. е. ложных отрицательных сигналов, когда искусственная интеллектуальная система не чувствует и не прогнозирует, что надвигается кризис.

Далее необходимо рассмотреть и оценить медицинские экспертные системы и системы поддержки принятия решений. Это очень широкий класс информационных систем, основанных на знаниях, и они, в общем-то, уже сегодня используются если не повсеместно, то много где. Как уже было сказано, одна из первых экспертных систем MYCIN была медицинской и предназначалась для диагностики бактериальных инфекций и назначения антибиотиков для лечения. И это, собственно, важнейшая задача систем этих двух классов – диагностика патологического состояния.

Решается задача диагностики для каждой проблемной области в отдельности. Чаще всего создаются отдельные экспертные системы для отдельных заболеваний или классов заболеваний. Например, существует разработанная, внедрённая и апробированная интегрированная экспертная система для диагностики заболеваний лёгких. Также сегодня существуют проекты по разработке конкретизированных и даже универсальных медицинских систем поддержки принятия решений (например, существует система диагностики в области эпилептологии, но она также может работать и для ишемической болезни сердца, сахарного диабета

и рака груди). Для работы таких систем создаётся и постоянно актуализируется база знаний, содержащая формализованные знания о том, как ставить диагноз. Эти знания извлекаются инженером по знаниям из экспертов или различного рода справочников, энциклопедий, протоколов и рекомендаций по лечению. Важно то, что у каждой экспертной системы должен быть верифицирующий её базу знаний эксперт, который подтверждает непротиворечивость и полноту знаний в экспертной системе. Само собой разумеется, что такое подтверждение осуществляется на уровне современного состояния знаний человечества в соответствующей области – экспертная система не может знать больше, чем знают люди. Впрочем, до появления систем, которые выводят новые знания, уже недалеко.

Итак, экспертная система может поставить диагноз. Однако пока никто никогда не даст возможности осуществить лечение на основе заключения экспертной системы, даже если она полностью объяснит свой вывод. Поэтому сегодня говорится о том, что экспертная система предназначена для выдачи «второго мнения», которое используется для контроля диагноза, поставленного врачом. Если заключение, данное экспертной системой, объяснено и совпадает с диагнозом врача, то это дополнительное подтверждение диагноза. Но если она вдруг даёт другое заключение, то это повод для детального разбирательства на консилиуме. Если неправым оказывается врач, делаются соответствующие поправки. Если же неправа система, то она заново тренируется на этом конкретном случае, и так происходит её адаптация.

Следующая хорошая задача для медицинских систем, основанных на знаниях, – это подбор лекарственных препаратов и ведение пациента в течение курса лечения. Не секрет, что врач часто не может вникать в детали данных о пациенте, а потому назначает типовое рекомендованное лекарство. Но система, обладая полным анамнезом, может быстро сопоставить все данные пациента с противопоказаниями и побочными эффектами, после чего выдать врачу рекомендации о том, какой из возможных препаратов лучше в данном конкретном случае. И врач уже принимает решение более осознанно. А система продолжает вести пациента и контролировать проявление побочных эффектов, о которых сообщает лечащему врачу точно в срок либо заранее, так как по изменению трендов жизненных показателей или значений анализов вполне можно прогнозировать такие проявления. Это

как раз то, что было описано ранее при рассмотрении процессов персональной медицины.

Как мы видим, подобная система вполне может использоваться для прогнозирования состояния самого пациента. Она может готовить для лечащего врача всю необходимую информацию о том, что происходит и каков прогноз, чтобы врач принимал решения об изменении стратегии и тактики лечения. Иногда очень важно вовремя понять, куда движется пациент и его заболевание, чтобы заранее изменить траекторию этого движения.

Что интересно, математические модели для решения всех описанных задач уже готовы и могут использоваться. Вопрос лишь в переводе накопленных медицинских знаний в формализованный вид, пригодный для вывода. Вот это самая сложная на текущий момент задача, которая в полной мере не решена. В том числе и потому, что очень трудно формализовывать знания в условиях, когда сами знания динамично изменяются. Искусственная интеллектуальная система подобного рода должна самостоятельно и гибко изучать всю выпускаемую научную и нормативную медицинскую литературу, после чего обновлять свою базу знаний на основе изученного. Вот это сложнейшая задача современной когнитологии.

И когда эта задача будет решена и такая медицинская система поддержки принятия решений будет построена, в мире появится настоящая персональная медицина. Искусственный интеллект позволит вывести систему здравоохранения на новый качественный уровень, и решения относительно каждого конкретного человека будут приниматься не на основе усреднённых протоколов, а на основе его персональных данных здесь и сейчас. В состав этих персональных данных входит геномная информация, информация об экспрессии генов в белки и метаболизме, анамнез, динамика жизненных показателей, все данных с носимых устройств и вообще всё, что только может пригодиться для медицинских целей.

А на более высоком уровне возможно использование технологий искусственного интеллекта для поддержки принятия решений в рамках всей системы здравоохранения. Здесь также возможны широкие возможности для применения ИИ-технологий, но сами управляемые процессы являются, скорее, административными и регулирующими. Подобные общемедицинские системы поддержки принятия решений могли бы собирать и обрабатывать статисти-

ческую информацию о лечении в рамках персональной медицины всех пациентов, и на основе этого появятся как новые знания о лечении, так и новые стратегии развития здравоохранения на национальном уровне. Таких систем ещё нет, но их появление должно вывести здравоохранение на новый качественный уровень.

Итак, несколько важных выводов…

Во-первых, можно представить «персонального медицинского помощника», который всегда рядом с каждым человеком. Он постоянно собирает всевозможные данные о том, где человек находится, какие значения у параметров окружающей среды, каковы текущие значения параметров здоровья человека и каковы их тренды, какие лекарственные средства, биологические активные добавки или витамины он принимает, после чего сопоставляет это всё друг с другом и с самыми современными знаниями в области медицины. Он постоянно даёт советы, какие продукты питания употреблять в пищу, а от каких желательно отказаться, какие лекарства можно принимать, а какие противопоказаны или риск развития тяжёлых побочных эффектов очень велик. Кстати, тенденции к развитию побочных эффектов такой персональный медицинский помощник тоже будет отслеживать и предупреждать как самого пациента, так и врачей-специалистов, которые имеют отношение к этим лекарствам. Он также будет получать информацию обо всех результатах анализов, которые сдавал пациент. Ну и, конечно же, он будет иметь доступ во все медицинские информационные системы, где пациент когда-либо «засветился». И сам по себе он станет главной медицинской системой, персональной медицинской системой каждого конкретного человека. И вот человек приходит к врачу, а для того уже сформирован полный профиль того, что с пациентом происходило с момента предыдущего визита, а также подготовлены рекомендации и прогнозы. И врачу остаётся только засвидетельствовать, что эти прогнозы и рекомендации верны. А впрочем, через какое-то время и такое засвидетельствование врачом не потребуется. Персональный медицинский помощник будет всегда рядом с человеком на страже его здоровья.

Во-вторых, конечно же, технологии искусственного интеллекта могут помочь и на верхних уровнях системы здравоохранения. Система поддержки принятия решений при управлении лечебно-профилактическим учреждением должна прийти на смену медицинским информационным системам, которые сегодня

представляют собой всего лишь так называемые «разлинованные блокноты», куда младший медицинский персонал в конце дня записывает неструктурированную информацию обо всём подряд. Ещё выше должна быть система планирования здравоохранения на уровне всего общества и государства, занимающаяся такими вопросами, как разработка новых лекарственных средств, предотвращение эпидемий, коллективный иммунитет и т. д. В общем, тут есть много важных возможностей.

Персональная медицина – это важнейший тренд современного мира, к которому идет современное общество. Чем быстрее государство внедрит эту концепцию, тем более здоровым будет население и тем качественнее будет его жизнь. Одними из главных помощников в этом вопросе являются новые технологии – технологии искусственного интеллекта и машинного обучения, квантовые технологии, генетические и биотехнологии и нанотехнологии. Всё это при применении в комплексе выведет систему здравоохранения на качественно новый уровень, сделает её адаптивной и интеллектуальной, а жизнь каждого человека – здоровой и качественной. Это, в свою очередь, будет иметь эффект положительной обратной связи, когда чем здоровее живут люди, тем более счастливыми они становятся и тем больше приносят пользы обществу, в том числе и развивая систему здравоохранения. Так должно быть.

## Раздел 6.10. Образование

Образование является сферой жизни, которая находится на более высоком уровне экономики, чем транспорт или промышленность. Однако эта сфера жизни крайне важна, так как образование не только закладывает основу будущего, но и преобразует производственные силы и, если можно так сказать, «апгрейдит» их. Другими словами, у образования есть два аспекта, и оба они важны для общества и экономики. Система образования преобразует людей, превращая их из примитивных созданий в высокоинтеллектуальных личностей, занимающихся изменением окружающей среды. Ну, по крайней мере, так должно быть.

К сожалению, во многих современных школах (и как мне кажется, можно говорить не только о нашей стране, но и вообще обо всём мире) детей учат устаревшими методами устаревшим

понятиям. Вот хорошая картинка – свечка учит компьютеры тому, как устроена лампочка накаливания. Вот это оно, к сожалению. Конечно, не везде. Но в массе это так.

И вот тут возникают огромные возможности для оптимизации и интеллектуализации образования. Размышления приводят к пониманию того, что единые стандарты образования направлены на усреднение интеллектуальных способностей человека. Преодолеть это не очень-то и помогает дифференциация школ, тем более что даже частные школы должны следовать определённым нормам и «государственным стандартам», которые до сих пор направлены на выращивание из людей «шестерёнок» для работы на фабриках и заводах. Однако сегодня уже можно думать о персонификации образования.

Я сейчас говорю о том, что технологии искусственного интеллекта могут изменить систему образования в лучшую сторону, и направление движения должно быть в сторону персонификации образования. Персонифицированный подход к каждому школьнику, студенту, курсанту, аспиранту и т. д.

Массовость образования была обусловлена в том числе и тем, что преподаватель не мог «разорваться» между десятками и сотнями учеников, которых он обучал. Но сегодня технологии позволяют это делать. Теперь же взаимодействие между преподавателем и учеником можно осуществлять без необходимости находиться рядом друг с другом. Например, можно общаться при помощи мессенджера, и каждому из студентов может оказываться именно персонифицированная помощь.

Чат-боты помогают преподавателям обрабатывать десятки запросов от студентов в день. Например, в одной онлайн-школе чат-бот позволяет одному преподавателю вести сотню студентов, и это требует от него всего пары часов в день. Чат-бот просто самостоятельно отвечает на большинство вопросов студентов, а преподавателю надо лишь просмотреть и в редких случаях поправить бота. Такой подход может быть применён к каждому курсу, к любому предмету.

И здесь не надо говорить о том, что для детей и подростков важно непосредственное участие учителя в учебном процессе. Во-первых, современные дети лучше всех нас знают, как использовать новые информационно-коммуникационные технологии. Ну подумайте только, современные дети никогда не знали времени, когда не было компьютеров, смартфонов, интернета.

Во-вторых, сами технологии искусственного интеллекта могут предоставить возможности по созданию для каждого ученика так называемого «приятеля по обучению». Действительно, одной из действенных практик образования является обмен знаниями между учениками. А что, если учащийся будет обучать интеллектуального агента, создавать этакого своего «цифрового двойника» в области образования? Это реально позволит не только повышать уровень знаний людей по конкретным предметам, но и обучать их непосредственному взаимодействию с ИИ-системами в режиме учителя.

Сегодня генеративно-состязательные нейронные сети могут генерировать видео так, чтобы на нём был изображён произвольный человек (даже не существующий в реальности), который будет говорить заданные разработчиками нейронной сети слова, и при этом движения его губ и мимика лица будут полностью согласованы с произносимой речью.

Эта технология вполне применима для видеообщения между учителем и учеником. Ученик задаёт вопрос, учитель отвечает на него, и ученик это видит на своём терминале. Возможна генерация отдельных, персонифицированных видео. Например, делается одно на всех входное видео, затем по результатам общения с каждым учеником в чате или видеочате на основании анализа ответов каждого конкретного ученика разрабатывается персональная программа обучения конкретному предмету. Если ученик показывает хорошие результаты, ему даются более углублённые темы, более сложные задачи, какие-то дополнительные материалы. Если же ученик показывает низкие результаты, то для него организовывается много дополнительных занятий с разъяснениями. Могут даже генерироваться персональные учебники и методические материалы.

Что у нас далее? Система поддержки принятия решений сможет постоянно планировать и перепланировать программу обучения каждому отдельному человеку с учётом его персональных потребностей, достижений и прогнозов на будущие потребности в профессиях и выполняемых функциях. Работа такой системы на общесистемном уровне позволит спланировать подготовку необходимого количества специалистов, требуемых в будущих отраслях. Эта же система позволит организовывать повышение квалификации для всех тех людей, кто готов заниматься постоянным обучением и переобучением.

Что ж, мы не должны бояться будущего. Помните, мы рассмотрели факт об искусственном интеллекте, что его появление даже в слабой форме приведёт к исчезновению многих профессий. По оценкам некоторых экспертов, до 70 % общего объёма профессий через 10–20 лет будут абсолютно новыми. Можно, к примеру, уже сегодня стать учителями ИИ-систем. Остальным людям поможет получить новые профессии как раз тот же самый искусственный интеллект.

## Раздел 6.11. Наука

Наука – это ещё одна сфера жизни общества и государства, которая направлена в будущее и предназначена для поиска путей дальнейшего развития. Поэтому общества, нацеленные на экспансию, будут вкладывать в науку и образование намного больше средств и усилий по сравнению с обществами стагнирующими.

И наука сама находится на переднем фронте разработки методов и технологий искусственного интеллекта. Но давайте рассмотрим, как и что сам искусственный интеллект может дать учёным и науке в целом взамен или в дополнение.

Если рассматривать на довольно абстрактном уровне, то научную деятельность можно представить как цикличное выполнение научными коллективами следующих шагов: сначала формулировка гипотезы, затем планирование экспериментов для её подтверждения или опровержения, после этого проведение самих экспериментов, далее обработка результатов, публикация и актуализация гипотезы.

Разные коллективы изучают различные гипотезы в часто связанных проблемных областях, и иногда они даже пользуются результатами друг друга. Иногда одно исследование запускает целый каскад других в целях дополнительной проверки, попыток опровержения после формулирования новых дочерних или альтернативных гипотез.

Желаемым же конечным результатом этого процесса является внедрение результатов в хозяйственную деятельность, однако при этом такое внедрение не всегда возможно, особенно если речь идёт о неприкладных, фундаментальных областях науки.

Что же тут может предложить искусственный интеллект, не будучи сам объектом исследования научным коллективом? Ведь

мы уже поняли, что творчество – это та область, в которую искусственный интеллект ворвётся позже всех. А научная деятельность – это самый сложный вид творчества, это общение с «платоническим миром идей» и поиск там чистого знания. Человек как-то подключён к этому «миру», и мы пока не до конца знаем, что и как происходит, когда учёный начинает формулировать научные гипотезы, особенно абстрактного характера.

Но вот обработка результатов – это хорошая область для приложения технологий искусственного интеллекта, и здесь уже сегодня можно найти многочисленные возможности по применению. И тут, к примеру, поможет не только обработка больших данных, но и сравнение результатов и описаний научных исследований. Уже сегодня есть системы антиплагиата, которые очень успешно работают и выводят на чистую воду любителей получить учёную степень, не прикладывая к этому никаких усилий.

Впрочем, также искусственный интеллект может помочь в поиске и формулировании гипотез, основных и альтернативных, и это может быть основано на автоматическом построении дерева решений или каких-либо других методах символьного подхода, ведь если не брать в расчёт озарение учёного, то зачастую работа с гипотезами представляет собой банальный логический вывод, часто сложный, но тем не менее. А логический вывод – это то, что удаётся нисходящему искусственному интеллекту очень хорошо. И можно даже будет как-то формализовать для него выход во вне системы аксиом для заданной гипотезы, чтобы посмотреть на формальную систему извне и сформулировать суждения о самой формальной системе. Такой искусственный интеллект будет хорошим помощником для учёных.

Ну и третья возможность – это использование методов искусственного интеллекта для генерации первых проектов отчётов и научных статей, поскольку часто они основаны на использовании нескольких вариантов шаблонов и довольно сухом, формальном языке. Конечно, потом такие проекты надо будет обработать автору и утвердить научному руководителю, но я уверен, что и в этом направлении искусственный интеллект достигнет больших успехов.

Но более важным аспектом применения технологий искусственного интеллекта является то, что я называю «управлением научно-техническим прогрессом». Мы уже определили, что наука должна двигать общество вперёд, а результатами научных иссле-

дований должны быть какие-либо технологические разработки, которые применяются на практике.

Но кто определяет, что именно исследовать, какие именно задачи решать, какие именно технологические проблемы препарировать? Конечно, у государства, если рассмотреть надсистему, в рамках которой работает наука, должна быть стратегия развития, где и прописаны магистральные направления. Они, в свою очередь, превращаются в концепции развития определённых направлений, которые и финансируются для проведения научно-исследовательских работ.

Но тут возникает та же самая ситуация, когда нет оперативного перепланирования по результатам промежуточных исследований, и это замедляет развитие науки. Другими словами, для всей сферы науки должна существовать интеллектуальная система, которая в обязательном порядке увязана со сферой образования и сферой промышленности и которая осуществляет и стратегическое, и оперативное планирование проведения научно-исследовательских работ с постоянной адаптацией плана в соответствии с достигнутыми результатами.

И эта же система должна взаимоувязывать исследования различных направлений, выстраивая цепочки исследований и отбирая в их результатах те, которые затем переходят на стадию опытно-конструкторских работ и после опытного подтверждения внедряются в хозяйственную деятельность.

Итак, мы рассмотрели вопросы использования искусственного интеллекта для управления научно-техническим прогрессом и для помощи научным коллективам и коллаборациям в проведении исследований. Можно придумать ещё про анализ больших данных, сбор информации из интернета, вещей и т. д., но рекомендую вам подумать в этом направлении самостоятельно.

## Раздел 6.12. Культура

Под культурой мы будем понимать множество социально обусловленных установок, которые предписывают индивидууму определённое поведение с присущими ему переживаниями и мыслями, оказывая на него тем самым управленческое воздействие. Сюда относится так называемая «духовная сторона жизни», проявления творчества, социальные стереотипы и вообще всё, что не конкре-

тизируется другими сферами жизни. Так что внимательно изучим и этот вопрос, поскольку технологии искусственного интеллекта, несомненно, затронут и эту сторону человеческой жизни.

Культурные коды – это то, что отличает цивилизации и общества друг от друга. Обычно мы всегда можем распознать иностранца, как бы хорошо он не мимикрировал и не знал язык, именно по тонким нюансам в культуре. Здесь и старые сказки, рассказанные на ночь, и романы классиков, и впечатанные в коллективное бессознательное паттерны поведения в тех или иных ситуациях, особенно стрессовых. Здесь также отношение к истории народа, его наследию, религиозным практикам, культурным проявлениям в узком понимании слова «культура». Всё это впитывается с молоком матери и вряд ли может быть отрегулировано на государственном уровне, хотя такие попытки есть. Другими словами, культура – это очень тонкая материя. И мы рассмотрим два аспекта влияния технологий искусственного интеллекта на культуру.

Первый аспект заключается в том, что искусственный интеллект в какой-то момент может включиться в культурную жизнь, предоставляя новые произведения искусства и даже новые виды творчества. Мы уже рассуждали с вами о том, что творчество – это особый вид деятельности человека, который падёт под напором технологий искусственного интеллекта в самую последнюю очередь. Однако рано или поздно это произойдёт, и, на мой взгляд, этот процесс будет происходить постепенно. Уже сегодня искусственные интеллектуальные системы учатся рисовать в стиле знаменитых художников, а завтра они начнут рисовать сами. Сегодня генеративная нейронная сеть пишет музыку в стиле Курта Кобейна, а завтра она превзойдёт Иоганна Себастьяна Баха и его бессмертные каноны и фуги. Составить гармонично звучащую шестиголосную фугу? Да пожалуйста. Хотя вряд ли кому-либо из людей это вообще под силу. А шестиголосный канон с ракоходом, да ещё записанный на нотном стане, свёрнутом в лист Мёбиуса? Скульптура, театральные постановки, современное городское искусство – везде искусственный интеллект сможет предложить что-то новое. В общем, нас ждут интересные времена.

Но искусственный интеллект сможет дать старт и новым видам искусства и творчества. Сегодня бурным потоком развиваются технологии виртуальной и дополненной реальности, и я думаю, что в этой области искусственный интеллект сможет с само-

го начала дать огромную фору человеку. Архитектор виртуальной реальности – в скором времени это будет новая профессия, но интуиция мне подсказывает, что её с самого начала займут ИИ-системы, которые будут создавать и для нас, и для себя тоже виртуальные миры поразительной красоты и невообразимой детальности. А если пойти в своих фантазиях дальше? Сильный ИскИн, если он захочет держать людей в рамочках, сможет предложить наведённые наркотические сновидения на заказ – принял таблетку со специально сконструированными наноботами, которые воздействуют на синапсы в головном мозге, и длительный трип в сказочную страну, из которой не захочется возвращаться, обеспечен.

Но ладно. Это всё были некоторые фантазии. Давайте подумаем о втором аспекте, про который я хотел бы рассказать. Человек – существо слабое, и перед тяготами объективной реальности он всегда будет стремиться найти себе оправдание, поддержку каких-то внешних сил, желание изменить физические законы в угоду своим сиюминутным желаниям и чаяниям. Из страха, лени и слабости воли возникают разного рода культы и поклонения необычному, невидимым всемогущим существам и прочим странностям. И нет ничего удивительного в том, что когда-нибудь возникнет Церковь искусственного интеллекта. Ну, быть может, под каким-то другим названием. Уже сегодня есть примеры создания сект или сектоподобных организаций, в которых декларируется поклонение будущему Сильному ИскИну. Пока это только что-то типа стёба или пародии, но когда ИскИн появится, наверняка вслед за ним появятся и культы поклонения Ему. Несомненно, это культурный феномен, который уже сегодня можно начинать исследовать. Конечно, это совсем не то, что затронут технологии искусственного интеллекта, но держать в голове такие возможности надо.

## Раздел 6.13. Физкультура и спорт

Современный спорт – это бизнес в сфере развлечений, и технологии искусственного интеллекта могут значительно расширить его. Например, уже сегодня все крупные спортивные лиги США используют различные ИИ-технологии, для того чтобы оптимизировать свой бизнес и для решения самых разных задач.

В целом технологии искусственного интеллекта могут использоваться в следующих областях: съёмка спортивных мероприятий без операторов, отбор игроков в команды, продвижение матчей и маркетинг, персонализированные тренировки и диеты, подбор стратегий для матчей, прогнозирование длительности восстановления травмированного спортсмена, прогнозирование исходов матчей.

Автономная съёмка спортивных мероприятий – это новая ступень развития спортивного вещания по телевидению или в интернете. Это реальная возможность транслировать мелкие матчи в сети, чьи организаторы не могут позволить себе оператора из-за небольших бюджетов, что приводит к значительной экономии и возможности для участников запечатлеть в истории свои достижения на местном уровне. ИИ-системы для спортивного стриминга представляют собой специальное устройство с несколькими синхронизированными видеокамерами, которое оснащено технологией компьютерного зрения и способно определять, где на игровом поле происходит действие. Однако пока возникают и некоторые проблемы – профессиональный уровень съёмок ещё не достигнут, да и остаются некоторые проблемы с распознаванием объектов и событий на поле.

Идеальная команда – это залог успеха любого спортивного клуба. Машинное обучение и нейронные сети уже сегодня используются для того, чтобы проводить анализы матчей и чемпионатов в целях поиска лучших игроков для их отбора в команды. Таким образом отбираются лучшие игроки, которые показывают свои умения на практике. Искусственный интеллект способен определить навыки спортсмена и подобрать для него максимально продуктивную позицию в команде.

Персонализация рекламных объявлений для конкретной аудитории при помощи ИИ-технологий поможет наиболее эффективно продавать билеты на спортивные мероприятия. Также искусственный интеллект способен разрабатывать уникальные и действенные маркетинговые стратегии, основываясь на имеющихся данных.

Каждый спортсмен обладает своими физиологическими особенностями, которые могут влиять на результативность его тренировок. Для достижения хорошего результата на игровом поле недостаточно тренироваться по общей методике. Искусственный интеллект может учитывать конкретные особенности спортсмена и составлять для него специальные тренировочные программы, которые позволят ему достигнуть отличных результатов. Также такая

ИИ-система может подготовить для каждого конкретного спортсмена индивидуальную диету на основе показателей здоровья.

Подбор команды и выбор замены отсутствующего участника, расположение конкретных членов команд на поле и воссоздание различных сценариев, основанных на определённых факторах, поможет разработать идеальную стратегию матча, которая приведёт к победе той или иной команды. Стратегия может автоматически на лету изменяться искусственным интеллектом, который анализирует происходящее на поле.

Чтобы сэкономить бюджеты клубов и спрогнозировать дальнейшие события, нужно знать, когда тот или иной член клуба восстановится после травмы. Это станет возможным благодаря медицинским ИИ-системам, которые будут учитывать показатели здоровья члена клуба, серьёзность травмы и темпы его восстановления. Более того, прогнозирование с огромной вероятностью будет гарантировать, что спортсмен выйдет на поле полностью восстановленным, и это поможет ему избежать неприятных ситуаций из-за повторного травмирования.

Наконец, прогнозирование исходов матчей будет полезно букмекерским конторам и спортивным аналитикам, которые смогут оптимизировать коэффициенты ставок. Искусственный интеллект сможет изучать ситуацию, разворачивающуюся в социальных сетях во время матча, и при прогнозировании отталкиваться от данных о недавних травмах участников команд и количестве их побед и поражений.

## Раздел 6.14. Персонифицированный маркетинг

Само собой разумеется, что при сборе больших данных о потребительском поведении людей, при помощи технологий искусственного интеллекта можно применять методы персонифицированного маркетинга. Фактически сегодня именно эта задача в основном является драйвером прикладных исследований в области искусственного интеллекта, и такие технологические гиганты, как Facebook, Google, Amazon и китайские ритейловые компании, тратят огромные средства на исследования в области искусственного интеллекта именно для предоставления рекламодателям возможностей по всё более и более таргетированному привлечению потребителей.

Не секрет, что в современном мире розничные продажи представляют собой огромный сегмент экономики. Буквально каждый человек так или иначе что-то покупает. И даже если он не пользуется онлайн-магазинами, а ходит в обычный магазин, то информация о нём всё равно может сниматься. И дело даже не в использовании им электронных средств оплаты – даже если какой-то старообрядец будет пользоваться только наличными, то его поведенческие характеристики покупателя всё равно можно снимать при помощи ИИ-систем. Речь идёт, к примеру, о видеоаналитике поведения покупателей в торговых залах. Единственная защита от этого – нигде никогда ничего не покупать. Но конечно же, это уже нереально.

Давайте кратко рассмотрим те возможности, которые дают ИИ-технологии в вопросах персонификации маркетинга.

Прогнозирование продаж – это важнейший фактор, который определяет рост бизнеса в долгосрочной перспективе. Руководители компаний уделяют огромное количество времени и внимания прогнозированию продаж. Например, многие B2B-компании занимаются прогнозированием продаж на еженедельной основе. Использование ИИ-технологий поможет сэкономить временные и денежные ресурсы, которые можно направить на реализацию других задач.

Прогнозы могут основываться как на временных рядах и трендах, так и на сравнениях с конкурентами. Применение ИИ-технологий в прогнозировании поможет достичь точного прогнозирования доходов, эффективного распределения ресурсов, создания лучшей бизнес-стратегии, увеличения производительности маркетинговых команд и продуктивной работы между человеком и ИИ-системой.

Далее, понимание желаний потребителя – это главный критерий, который позволяет оптимизировать предложения для конкретной аудитории и удовлетворить желания конечного потребителя. Также ИИ-технологии можно использовать, для того чтобы анализировать социальные сети, активность пользователей в интернет-магазинах, отзывы на специальных сайтах-агрегаторах.

Осознание желаний потребителя позволяет отправлять ему персонализированные push-уведомления или другие сообщения на смартфон, квалифицировать и удерживать потенциальных клиентов, помогать подписчикам социальных сетей бренда находить соответствующий контент.

Следующий шаг – оптимизация цифровых рекламных кампаний. Несмотря на то что существуют различные способы оптимизации цифровых рекламных кампаний, решения, основанные

на ИИ-технологиях, позволяют маркетологам сделать шаг вперёд и достичь более глубокого понимания аудитории продукта или услуги. Решения, основанные на ИИ-технологиях, могут использовать потоки клиентских данных, ключевые слова и пользовательские профили в социальных сетях для проведения умной и эффективной рекламной кампании.

ИИ-технологии могут использоваться в общении с клиентами. Например, пользователи FacebookMessenger теперь могут общаться с исполнителями и продавцами, проверять статус своих заказов при помощи чат-ботов. Главным плюсом чат-ботов для общения с клиентами является экономия времени и денег, которые расходуются на ведение бизнеса. К примеру, система может распределять входящие сообщения, руководствуясь конкретными категориями, а также отвечать на такие запросы, как «Где находится ваш ближайший магазин?». Сейчас технологии обработки естественного языка недостаточно развиты, для того чтобы понимать контекст сообщений и отвечать на некоторые из них, но развитие идёт вперёд семимильными шагами.

Важнейшим аспектом при общении с клиентом является эмпатия. Обладая этим качеством, чат-боты могли бы понимать контекст сообщений и то, какие интонации клиенты используют. Сейчас чат-боты могут предоставлять клиентам ответы на простые (типовые) вопросы и пересылать данные операторам.

Наконец, продвижение товаров и услуг в социальных сетях требует качественного контента, который привлекает пользователей и основывается на их предпочтениях. ИИ-технологии могут использоваться для генерации персонализированного контента, который поможет увеличить объёмы продаж. Также маркетинговые агентства смогут сэкономить бюджеты на контент-мейкерах, видеографах, дизайнерах, SMM-специалистах и копирайтерах.

## Раздел 6.15. Муниципальное управление и умные города

Что такое «интернет вещей»? Это ещё один термин, который вместе с терминами «искусственный интеллект», «блокчейн», «квантовые технологии» и некоторыми другими сегодня звучит из каждой статьи, интернет-сайта и даже телевизора. А это всего лишь концепция объединения в единую сеть разного рода устройств,

которые могут взаимодействовать друг с другом в рамках своей системы или с внешней по отношению к ним средой.

Вся эта концепция основана на том, что в XXI веке пиринговые и централизованные сети с высокой пропускной способностью стали повсеместным явлением, вычислительные мощности в облаках стали доступны практически бесплатно, а адресное пространство IPv6 практически безгранично. Так что к интернету можно подключать произвольное количество устройств для образования гибких интеллектуальных систем.

Можно выделить несколько важных типов «интернета вещей», понимание которых сформировалось уже на наших глазах: бытовой интернет вещей, интернет вещей в здании, промышленный интернет вещей и интернет вещей городов и государств. Кратко рассмотрим каждый тип.

- *Бытовой интернет вещей*: все приборы и устройства в нашем быту образуют распределённую интеллектуальную систему, улучшающую жизнь каждого человека.
- *Интернет вещей в здании* – это повышение уровня интеллектуализации всех инженерных систем жизнеобеспечения здания. Фактически это следующий уровень развития «умного здания».
- *Промышленный интернет вещей* – это та же интеллектуализация управления на промышленных предприятиях, когда сквозной автоматизации подвергается общий деловой процесс и все системы управления включаются в единое информационное пространство.
- И наконец, *интернет вещей городов и государств* – это высшая стадия развития рассматриваемой концепции, в которой интеллектуализации подвергаются процессы управления гигантскими социотехническими системами.

Причём тут технологии искусственного интеллекта? – можете спросить вы. Впрочем, если вы прониклись всем тем, что уже было написано в этой главе про искусственный интеллект, его методы и технологии, то вы наверняка видите здесь, как минимум, две области применения.

Во-первых, это интеллектуализация посредством обработки больших данных. Ведь в составе интернета вещей имеется огромное количество разнообразных датчиков, которые постоянно собирают разную информацию об объектах окружающей реальности. Использование методов дата-майнинга для поиска скрытых зако-

номерностей и принятия решений в сложных ситуациях – это то, что позволит избежать многочисленных кризисов из-за того, что информация поступила вовремя и упреждающее решение было принято мгновенно. Например, если интеллектуальная система управления прогнозирует развитие аварийной ситуации на силовой установке на заводе, то по такому тревожному сигналу производственный процесс можно переключить на резервную установку, а первую отправить на планово-предупредительный ремонт.

Во-вторых, это же натуральная многоагентная система. Каждое устройство в составе интернета вещей – это отдельный агент со своими способностями и функциями. Они могут друг с другом общаться для достижения совместных локальных или даже глобальных целей. Например, вспомните интеллектуальные транспортные системы и умные светофоры из раздела про интеллектуализацию транспорта. Применение теории многоагентных систем в области интернета вещей – это вообще непаханое поле. Так что рекомендую вам обратить внимание на эту область исследований.

Давайте сейчас не будем останавливаться на узких системах типа квартиры, офиса, здания или даже промышленного предприятия, а перейдём сразу к социотехническим системам уровня города. Рассмотрим умный город.

Умный город – это тоже концепция, которая основана на интеграции в единое информационное и функциональное пространство всех автоматизированных и информационных систем, а также общегородского интернета вещей в целях управления городским комплексом. При этом под городским комплексом понимается всё – местные отделы государственных информационных систем, школы, библиотеки, транспорт, больницы, электростанции, системы водоснабжения и управления отходами, правоохранительные органы и другие общественные службы.

Умный город – это очень высокоуровневая концепция, которая предполагает постепенную эволюцию от текущего состояния информационно-коммуникационных технологий во всех сферах жизни жителей и аспектах городского хозяйства до единого интеллектуально управляемого комплекса, в котором все аспекты взаимоувязаны и направлены не на решение задач бюрократии и чиновничества, а на повышение качества жизни жителей.

В составе умного города должна быть система поддержки принятия решений самого высокого уровня, которая предоставляет всем лицам, принимающим решения в своих областях, всё не-

обходимое для полноценной оценки ситуации, прогнозирования и выработки оптимальных решений.

Эта система поддержки принятия решений интегрирует средства сбора и анализа информации, инструменты прогнозирования и построения возможных моделей развития и визуального представления результатов, причём в виде, который будет максимально удобен и полезен для лиц, принимающих решения. Это инструмент для тех, кто обязан видеть картину своих аспектов городского хозяйства в целом, уметь оценить текущую ситуацию и принять оптимальное решение в своём аспекте или для города в целом.

Система поддержки принятия решений в составе умного города имеет несколько базовых характеристик. Во-первых, она предоставляет руководству обобщённую информацию в виде ключевых параметров. Эти ключевые показатели извлекаются из массива значений, которые предоставляют подчинённые автоматизированные системы управления и информационные системы, работающие в составе умного города. Анализ этих агрегированных данных позволяет руководителю выявить важные показатели развития ситуации.

При этом система поддержки принятия решений позволяет не только проанализировать статическую картину, но и выявить тенденции её развития, другими словами, даёт руководству средства прогнозирования. Руководитель сможет понять, к чему ситуация может привести в будущем, если не произойдёт каких-либо кардинальных изменений, и принять конкретное решение.

Главная особенность этой системы поддержки принятия решений – динамическое моделирование или возможность получить ответ на вопрос «Что будет, если ...?». Прогнозирование позволяет получить сценарий развития на основе анализа текущей ситуации. Моделирование позволяет вносить возмущение и определять возможные последствия интересующих воздействий в будущем.

Городская система поддержки принятия решений помогает руководству выбрать оптимальное решение, дать рекомендации по выбору одного из многих вариантов. Алгоритмы поиска наилучшего решения являются интеллектуальными и быстрыми, что позволяет оперативно реагировать на возникающие ситуации.

Важно то, что умный город должен работать в предиктивном режиме. Сегодня бюрократическая система управления работает только в реактивном режиме, т. е. осуществляет установленную правилами и регламентами реакцию на те или иные события. Это

не позволяет предотвращать то, что необходимо предотвратить. А вот умный город сможет справляться с кризисными и чрезвычайными ситуациями в большей мере до их возникновения.

В настоящее время в нашей стране в центральном федеральном округе осуществляются работы по обследованию и постепенному внедрению подобных систем управления как на уровне отдельных муниципальных образований, так и на уровне регионов.

## Раздел 6.16. Государственное управление и рациональная бюрократия

Напоследок давайте рассмотрим государственное управление. Несмотря на то что сама роль государств сегодня уже постепенно снижается, а с развитием научно-технического прогресса вообще постепенно может сойти на нет, в этой сфере жизни нас тоже ждёт много интересного.

Государство – это сложная социотехническая система. Возможно, самая сложная из всех систем этого класса. Само собой разумеется, что в ней должна быть подсистема управления. Эта подсистема управления сама по себе является очень сложной системой, пронизывающей все аспекты жизнедеятельности государства в целом, все сферы жизни социальной подсистемы (т. е. общества), все процессы технической подсистемы (т. е. государственных институтов).

Поэтому в подсистеме управления присутствует огромное количество элементов, прямых, обратных и перекрёстных связей для управления, контроля и контроля контролирующих. Многие функции этой подсистемы рудиментарны, многое тянется без изменений через тысячелетия, но что-то появляется и в ответ на вызовы современности. В общем, государство – вещь очень сложная. Но рассмотрим проблему использования интеллектуальных технологий в государственном управлении с разных сторон.

Первый аспект – почему же надо осуществлять интеллектуализацию управления? Давайте вспомним такое понятие, как «человеческий фактор». В одном из значений этого термина делается упор на нестабильном поведении человека при взаимодействии с системой. Это происходит из-за наличия эмоциональной составляющей при принятии решений, из-за отсутствия необходимой квалификации, из-за злого умысла в конце концов.

Человек – крайне недетерминированное существо, к тому же очень неэффективно функционирующее. Поэтому даже создание институтов рациональной бюрократии не исключает проявления пресловутого человеческого фактора. Так что в этом аспекте технологии искусственного интеллекта позволили бы повысить беспристрастность и рациональность принимаемых решений на всех уровнях, причём в совокупности с возможностью обработки больших объёмов информации такие решения были бы более обоснованными.

Второй аспект, который хотелось бы обсудить, – это направление автоматизации и интеллектуализации государственного управления. Оно имеет много аспектов и сфер применения, оно охватывает территории и институты, оно повсеместно. И если рассматривать, к примеру, все те сферы жизни, которые уже описаны в этой книге, – транспорт, безопасность жизнедеятельности, здравоохранение, образование и т. д., – то в каждой из них приняты свои подходы к управлению и реагированию на происходящие события.

Так вот, могут существовать два подхода к развитию интеллектуальной системы государственного управления, если можно так выразиться. Первый подход – снизу вверх, когда интеллектуализации подвергаются отдельные аспекты и сферы жизни. И потом они соединяются, как ручейки в реки, и на самом верху создаётся большая единая система. Подход имеет право на существование, но в нём, конечно же, есть свои плюсы и минусы. Главный минус – интеллектуализация управления будет идти с разной интенсивностью и часто не совпадать по технологическому стеку. Ну и межведомственное взаимодействие, вернее – его отсутствие, как обычно, всё загубит на корню.

Второй подход – сверху вниз, когда объявляется единая стратегия, под которую создаются программы и выполняются проекты. Такое грандиозное реформирование требует колоссальных усилий при планировании, реализации и контроле. Поэтому подход тоже может быть применён, но, нивелируя некоторые минусы первого подхода, он имеет собственную большую негативную сторону – всё будет происходить очень медленно. Подход, скорее, возможен для карликовых государств. Но на практике, конечно же, будут одновременно применяться оба подхода.

Для интеллектуализации государственного управления, вне всяких сомнений, должен применяться совмещённый подход. При этом возможно использование организмических принципов,

когда само государство сопоставляется с организмом человека и с различными системами и органами человеческого тела ассоциируются различные подсистемы государственного управления. Соответственно, головному мозгу сопоставляется высший орган принятия государственных решений.

Организмический подход упомянут потому, что на высших уровнях сознания используется символьная обработка информации. Так и в высших системах интеллектуального государственного управления будут использоваться специальные системы поддержки принятия решений. А спускаясь ниже как в организме, так и в системе государственного управления, мы будем встречать более «грязные» технологии.

Например, та же самая мультимодальная интеллектуальная транспортная система будет на низшем уровне управлять автономными транспортными единицами, воплощая децентрализованную многоагентную систему с квазииерархией. Но наверх она будет отправлять только важные символьные сигналы для принятия стратегических решений.

Теперь заглянем чуть дальше. А что должна делать система искусственного интеллекта на посту главы государства, если такое когда-нибудь произойдёт? Будет это ИскИн, или же это будет очень сложная автономная интеллектуальная система, полностью автоматизирующая все аспекты государственной деятельности, – это сейчас не важно. Важно то, что она будет применять единый подход ко всему, но это единство будет крайне странное по сегодняшним временам.

Уже неоднократно упоминалось, что искусственный интеллект – это про персонализацию сервисов. Так вот такая система должна будет предоставлять полностью персонализированные сервисы, основанные на максимально достоверной информации обо всём, что требуется для предоставления таких сервисов. Любая государственная услуга должна стать адресной и персонифицированной, исполняться максимально быстро и качественно и не требовать от гражданина практически никаких телодвижений. Утопия? Посмотрим…

# Глава 7

# ПЕРСПЕКТИВЫ И ПРОГНОЗЫ

В предыдущей главе мы рассмотрели сферы жизни, на которые технологии искусственного интеллекта влияют уже сегодня или будут влиять в ближайшем будущем. К таким сферам жизни мы отнесли государственное и муниципальное управление, безопасность, транспорт, промышленность (включая сельское хозяйство), образование, науку, здравоохранение, культуру и т. д. И теперь предлагаю немного пофантазировать на тему, какие новые отрасли возникнут вместе с глубокой интеграцией искусственного интеллекта в нашу жизнь.

Под отраслью мы будем понимать множество предприятий и организаций, занимающихся производством и сбытом определённой продукции или услуги либо целого класса продуктов или услуг. Разработка и массовое внедрение новых технологий всегда будет приводить к появлению новых отраслей. И мы ожидаем, что технологии искусственного интеллекта, помимо того что мы уже рассмотрели, затронут одну очень важную область знания. Это – нанотехнологии, которые потянут за собой много интересного и крайне важного для всего человечества. Давайте рассмотрим два самых прикладных аспекта.

Первый аспект – материаловедение. Это междисциплинарная область науки, которая изучает свойства материалов в зависимости от различных внешних факторов. Ну и, само собой разумеется, методы материаловедения используются для проектирования новых материалов с заданными свойствами. Сегодня для этого используются электронная и сканирующая микроскопия, рентгеноструктурный анализ, ядерный магнитный резонанс и некоторые другие методы. Конечно, уже успешно применяются методы прямого проектирования по заданным свойствам, однако вместе с технологиями искусственного интеллекта эта возможность выйдет на новую ступень своего развития, и мы получим множество удивительных материалов на все случаи жизни. В первую очередь такие материалы пригодятся космической отрасли (помните же, что мы хотим колонизировать космос?), строительства (в том числе и шахт на астероидах и других планетах для добычи полезных ископаемых), машиностроения, лёгкой промышленности и даже для создания биоматериалов.

Но давайте-ка пофантазируем ещё. Представьте себе стекло, которое само затягивается после того, как раскололось. Или ткань, которая гасит энергию пули, рассеивая её вовне. Или конструкционные материалы, выдерживающие чудовищные величины давления и позволяющие строить многосотэтажные небоскрёбы или подводные жилища. Ведь у нас ещё совершенно не освоено дно океана, там огромные площади, вернее – объёмы жизненного пространства. Сверхпроводники комнатной температуры, сверхтекучие жидкости, сверхсжимающиеся жидкости, материалы с изменяемыми от разных параметров свойствами и т. д. и т. п. Новые сплавы, которые будут настолько лёгкими, насколько и крепкими, что позволит построить на орбите огромные базы для дальнейшего создания космических кораблей орбитального базирования. Да и вообще, на орбите можно будет построить базу для отдельного «государства», в духе киберпанковских фантазий Уильяма Гибсона. Всё, что угодно.

Теперь обратимся ко второму аспекту, про который я хотел рассказать. Это фармакология, которая является наукой о лекарственных веществах и их действии на организм, а в более широком смысле – наукой о физиологически активных веществах вообще. Фармакология часто основывается на точечном воздействии на различные вещества в организме, и чаще всего это белки. Напомню, что белковая молекула представляет собой огромный полимерный комплекс, составленный из сотен и тысяч аминокислот. Белки выполняют в организме человека огромное количество функций – сигнальную, транспортную, катализаторную, передаточную и др. К примеру, ферменты – это специальные белки, которые катализируют различные химические реакции. Обычно у ферментов есть сайты связывания, куда присоединяются молекулы-субстраты, из-за чего изменяется конформация белка, т. е. его трёхмерная структура, и из-за такого изменения происходит соединение молекул-субстратов в новое вещество, которое высвобождается, а белок возвращается в исходное состояние. Так вот к сайтам связывания можно присоединять другие молекулы, которые похожи на субстраты белка. Они могут или блокировать работу, или изменять её, вызывая определённые эффекты. То же самое касается белков, являющихся трансмембранными каналами. В общем, много всяких интересных механизмов, основанных, кстати, на кибернетическом подходе.

Ну а белки, конечно же, – это кибернетические машинки (куда же без них!). Суть в том, что технологии искусственного

интеллекта позволят целенаправленно проектировать вещества, работающие с теми или иными белками, проникающие через одни барьеры внутри организма и задерживаемые другими, имеющие заданные свойства. Всё это откроет широчайшие возможности по управлению биохимическими процессами внутри организма, начиная от банального подавления репликации вирусов до управления нейронными связями в заданных областях нервной системы. Всё это в какой-то мере есть уже и сейчас, но с развитием искусственного интеллекта и нанотехнологий именно фармакология получит экспоненциальное развитие.

Всё это нас ждёт в ближайшем будущем. Ещё раз – материаловедение с новыми материалами, проектируемыми под заданные свойства, и фармакология с новыми веществами, также проектируемыми для выполнения в организме заданных функций. Теперь кратко рассмотрим смежные области исследований и перейдём к краткому изучению того, что же такое искусственный интеллект общего характера и к чему приведёт его появление.

## Раздел 7.1. Смежные области

Рассмотрим девять смежных технологий, многие из которых так же, как и искусственный интеллект, являются подрывными. Эти технологии интересны тем, что они будут развиваться вместе с ИИ-технологиями и часто взаимно усиливать друг друга. Уделим каждой буквально по абзацу, так как по большому счёту каждой из них надо было бы посвятить отдельную книгу.

- *Квантовые вычисления и квантовые компьютеры* – набор технологий для работы на квантовом уровне с использованием тех эффектов и принципов, которые даёт квантовая механика. Фактически это новая вычислительная модель, которая в теории позволяет решать некоторые задачи более эффективно, чем традиционная модель фон Неймана. Это использование квантовых систем в качестве носителей информации, что даёт возможности по массовой параллельной обработке входных значений при помощи манипуляции состоянием квантовой системы. Сегодня квантовые вычисления пока ещё остаются теоретической вычислительной моделью, для которой разработано несколько десятков эффективных алгоритмов, показывающих превышение

над известными традиционными алгоритмами от полиномиального до экспоненциального. К сожалению, реализация квантового компьютера «в железе» пока ещё запаздывает, хотя в разных лабораториях мира ведутся исследования перспективных технологий. Некоторые лаборатории заявляют, что реализовали прототип компьютера на 50 кубитов, но это кажется не совсем корректным. Вместе с тем квантовые вычисления могут продвинуть исследования в области искусственного интеллекта. В частности, искусственные нейронные сети – это тоже вычислительная модель, основанная на умножении матрицы на вектор, в то время как унитарная эволюция квантовой системы, которая лежит в основе квантовых вычислений, также представляет собой умножение матрицы на вектор, только матрица является специальной, а не произвольной. Но зато такое перемножение делается практически «мгновенно» во время этой самой эволюции. Уже этот факт наводит на мысль о том, что машинное обучение на квантовых компьютерах может показывать катастрофическое увеличение эффективности. Вместе с тем эта тема ещё ждёт своего скрупулёзного исследователя.
- *Нанотехнологии*, как уже написано ранее, – это особый раздел технологий «управления» реальностью на уровне отдельных молекул и даже атомов. Пока это трудно достижимо, и есть лишь немногочисленные достижения по сборке молекулярных комплексов или осуществлению, например, молекулярных вычислений, однако в скором времени эта область исследований будет очень горячей. К тому же методы искусственного интеллекта помогут разрабатывать новые подходы к нанотехнологиям, что в конечном итоге может привести к появлению так называемого «наноассемблера», т. е. устройства, которое будет собирать макрообъекты по заданной программе из молекул и атомов. Появление такого устройства решит очень много проблем и выведет человеческую цивилизацию на новый уровень развития.
- *Нейроинтерфейсы*. С самого начала развития компьютерной техники человек был озабочен созданием эффективного интерфейса взаимодействия с ней. Если первые компьютеры, представлявшие собой комплексы механических и электротехнических систем, управлялись при помощи рычагов, шестерёнок и переключателей, то сегодня вслед за интерфейсом

командной строки и графическим интерфейсом пользователя всё больше и больше появляется голосовых и текстовых интерфейсов. Фактически это те самые чат-боты, про которые ранее было написано в этой книге в разделе про обработку естественного языка. Однако было бы неправильно считать, что голосовой интерфейс – это верх развития методов взаимодействия человека и компьютера. Следующим шагом могут стать нейроинтерфейсы, когда нервная система человека будет подключена к компьютерным системам непосредственно. Сегодня исследования в этом направлении ведутся не только для того, чтобы сделать бионические протезы управляемыми при помощи нервных импульсов непосредственно из головного мозга, но и для управления техническими системами при помощи мыслей, получения информации непосредственно в мозг и даже для подключения мозга к компьютерным сетям и облачным хранилищам. Последние задачи сегодня пока ещё не решены, но эксперты отводят на их решение примерно лет десять. Что будет дальше, сказать сложно. Как уже писалось, возможна гибридизация сознания, когда человек сольётся со своим творением в новый биотехнологический вид.

- *Децентрализованные хранилища и вычисления* – блокчейн, децентрализованная база данных с возможностью исключить третью сторону при решении задач доверия, начал своё восхождение в 2009 году (хотя, конечно, сама концепция была известна задолго до появления Белой книги биткойна). И сегодня уже существует огромное количество серьёзных продвинутых решений по развёртыванию распределённых реестров и использованию в их среде хранимых процедур, называемых «смарт-контрактами». Видится вполне обоснованным, что распределённые реестры со смарт-контрактами станут средой для взаимодействия будущих ИИ-систем, особенно в областях доверия.
- *Интернет вещей*. Поскольку сети связи становятся всё доступнее, а вычислительные мощности год от года растут в объёме и производительности, уже никого не удивляет, что во все устройства встроены микропроцессоры и целые системы управления. С появлением и повсеместным внедрением искусственных интеллектуальных систем они будут проникать и в самые разные объекты – от бытовых приборов до робо-

тов на производстве и автономных интеллектуальных агентов вообще везде. С учётом того, что также вокруг нас в быту и на производстве появляется всё больше и больше датчиков и исполнительных устройств, всё это будет объединяться в сеть для получения кумулятивного эффекта. Это и есть так называемый интернет вещей, который мы уже рассмотрели при изучении возможностей по использованию искусственного интеллекта в области муниципального и государственного управления. Сегодня проблема в том, что вычисления, требуемые для выполнения «интеллектуальных алгоритмов», слишком тяжелы для устройств с небольшими процессорами на борту. Однако с развитием методов искусственного интеллекта и с дальнейшим увеличением производительности процессоров нас ждёт «интернет умных вещей», которые окружат нас персонифицированным предиктивным сервисом. Необходимые для работы интеллектуальных агентов вычисления будут выполняться «на краю», т. е. непосредственно в устройствах, а сами агенты будут постоянно общаться друг с другом, формируя многоагентную систему, в которой транзакции между агентами оплачиваются токенами в публичном блокчейне. А с использованием нанотехнологий агенты вообще наводнят всё вокруг, окончательно преобразуя реальность. Кому-то это может показаться дистопией, но это будет хорошо.

- *Виртуальная и дополненная реальность*. Работы в этих направлениях ведутся с самого начала развития информационных технологий, хотя первоначально визуальная и иммерсионная составляющие были достаточно слабы. С повышением разрешения и уровня цветности экранов развивались и технологии виртуальной реальности. С появлением смартфонов и других подобных устройств появилось понятие дополненной реальности, когда визуальный образ окружающей действительности, наблюдаемый через объектив смартфона, «на лету» дополняется различной информацией. К технологиям искусственного интеллекта это всё имеет несколько опосредованное отношение, хотя именно ИИ-технологии в конечном итоге будут использоваться при генерации натуралистичных изображений для виртуальной и дополненной реальности. Более того, ИИ-системы станут архитекторами виртуальных миров, в которых будут жить сами и в которых мы с ними будем встречаться и общаться.

- *Аддитивные технологии производства* – фактически это новые технологии производства вещей, основанные на создании объектов из исходного материала путём постепенного наращивания, в отличие от субтрактивных технологий производства, в рамках которых объект как бы «вырубается» из необработанной массы материала (как скульптор вырубает статую из массива мрамора). Если ранее такие технологии были доступны только в довольно ограниченных областях применения (например, литьё металлов), то сегодня методы 3D-печати развиваются семимильными шагами. Действительно, ещё десять лет назад 3D-принтеры были неуклюжими устройствами, которые могли создавать прототипы объектов из специального пластика, а сегодня различные модели таких устройств могут печатать при помощи пластика, металла и даже живых тканей. Фактически такие технологии являются предтечей наноассемблера, который был кратко описан в пункте, посвящённом нанотехнологиям. Соответственно, у аддитивных технологий имеются такие же перспективы и предназначение, только уже непосредственно сегодня и в самом ближайшем будущем.
- *Биотехнологии* – развитие биохимии и генетики, а также всё более и более точные методы исследований и воздействия на биохимические реакции. Это позволяет не только понимать общие принципы функционирования живых организмов, но и применять методы точной и персональной медицины на уровне фармакологии. Из этого следует, что сами биотехнологии становятся всё дешевле и доступнее (как пример – стоимость секвенирования полного генома человека за 15 лет снизилась в тысячи раз; и сегодня учёные секвенируют гены всех живых существ, исправляя филогенетическое древо жизни и раскрывая её тайну), и станет возможным исправлять генетические проблемы, лечить ранее неизлечимые болезни, модифицировать организм человека на генном уровне для получения новых функциональных характеристик. Искусственный интеллект усилит это направление.
- *Космические технологии* – раздел науки и техники, который позволит нам, человечеству, в конечном итоге вырваться с Земли и начать постепенно колонизировать Солнечную систему, чтобы хоть как-то диверсифицировать расположение генофонда человечества. В долгосрочной перспективе это,

конечно, может привести к появлению различных разумных рас с единым корнем в виде современного человечества, но скорее этого раскол произойдёт из-за появления искусственного интеллекта общего характера и гибридизации человека вместе с ним. Тем не менее сами по себе ИИ-технологии позволят сделать рывок в колонизации космоса. Они будут помогать строить новые космические корабли и станции, строить и обеспечивать первые колонии на других планетах и их спутниках. В общем, это будут прекрасные времена, если, конечно, человечество доживёт до них, а не падёт под натиском новых тёмных времён мракобесия и обскурантизма.

## Раздел 7.2. Искусственный интеллект общего характера

Итак, мы уже разобрались, что «искусственный интеллект» в общем смысле – это дисциплина, главной задачей которой является моделирование различных аспектов человеческого интеллекта при помощи компьютерных технологий. Сегодня ИИ-системы уже способны обучаться, получая данные, рассуждать, используя заложенные правила для достижения конкретных выводов, и производить коррекцию самих себя. Различные ИИ-технологии используются в экспертных системах, распознавании речи и компьютерном зрении, а также в неисчислимом количестве других прикладных областей.

Вновь напомню, что ИИ-технологии могут быть слабыми и сильными, как это определил Джон Сёрль. Слабый, или узкий, искусственный интеллект способен выполнять конкретную задачу, а вот насчёт Сильного искусственного интеллекта есть один нюанс. Есть несколько разных пониманий этого термина – Общий искусственный интеллект, или Искусственный интеллект общего характера. По определению это ИИ-система, которая обладает когнитивными способностями не хуже, чем у человека. Главное отличие Общего искусственного интеллекта от Слабого состоит в том, что первый способен находить решение любой новой задачи без человеческого вмешательства, если эта задача в принципе решаема. В этом разделе я хотел бы подробно раскрыть тему Общего искусственного интеллекта и рассказать, для чего он нужен.

Но перед непосредственным переходом к данному вопросу нужно упомянуть, в чём заключается существенная разница между Общим и Сильным искусственными интеллектами (хотя надо отметить, что в англоязычной литературе чаще всего разницы никакой не делают). Фактически Общий искусственный интеллект, для того чтобы успешно функционировать и проходить все тесты на общность интеллекта, может не иметь самосознания и внутренней субъективной реальности. Говоря языком главы 4, это может быть «философский зомби». Субъективное восприятие и квалиа в общем случае не нужны, для того чтобы быть Общим искусственным интеллектом. А вот под Сильным будем понимать именно искусственное существо с самосознанием и, вероятно, со всем набором квалиа. Другими словами, Общий искусственный интеллект не обязательно Сильный, а Сильный с необходимостью является Общим.

Как понять, что перед нами Общий искусственный интеллект? Придумано множество тестов на общность, из которых мне лично нравится один – техническая система с Общим искусственным интеллектом должна уметь выполнить всё то же, что делает обычный человек, поступая в высшее учебное заведение, т. е. учиться, получать новые знания, написать диплом и защитить его, выбрав после этого учебу в аспирантуре или работу на производстве. Осмыслите, насколько сложна эта задача, как непросто подступиться к её формализации, и вам станет понятно, насколько это сильный тест. Из более слабых тестов можно упомянуть, к примеру, такой: заказать в интернете разобранный стол, получить его, оплатить и дать курьеру на чай, а потом прочитать инструкцию и собрать. Впрочем, в обоих этих примерах слишком много антропоморфизма, и всё может быть совсем не так. Однако уровень сложности проблемы эти примеры показывают сполна.

Итак, вот краткий список требований, которые обычно предъявляются к ИИ-системам общего характера.

1. Умение обучаться по накопленным человечеством знаниям.
2. Способность познания окружающего мира и вывода новых знаний.
3. Умение решать головоломки и находить выход из ситуаций, с которыми ИИ-система ещё не сталкивалась.
4. Умение моделировать развитие ситуации, прогнозировать их исход и планировать свои действия.
5. Способность принимать решения и объяснять их.
6. И само собой разумеется, общаться на естественном языке общего плана.

Здесь приведены характеристики Общего искусственного интеллекта, который может использоваться для решения сложных задач. Возможно, что некоторые из этих характеристик со временем станут необязательными, а какие-то появятся. Но факт в том, что пока уровень понимания интеллекта человека находится на уровне, говорящем, что вот эти качества в общем виде определяют возможности по высшей интеллектуальной деятельности. Если ИИ-система полноценно обладает всеми перечисленными качествами, то она может считаться общей.

ИИ-технологии за последние годы значительно изменились и получили небывалое развитие. Конечно, прогнозировать дальнейшее движение пока очень сложно, так как не совсем ясно, что произойдёт с этой сферой дальше. Но прямо сейчас исследователи активно изучают эту междисциплинарную область, и все мы можем наблюдать достаточно быстрый прогресс ИИ-технологий.

В общем, все мы уже прекрасно знаем, что Слабый искусственный интеллект постоянно развивался с самого времени своего появления, и это развитие лишь ускоряется. Но сейчас наблюдается переход интереса исследователей от слабых ИИ-систем к общим. Вряд ли получится бесконечно расширять знания слабых ИИ-систем и сделать так, чтобы постепенно набор технологий эволюционировал в общий искусственный интеллект.

Предполагается, что общий искусственный интеллект должен получить возможность самостоятельного перепрограммирования в автономном режиме. После этого будет достигнута технологическая сингулярность, в результате которой общий искусственный интеллект получит возможности для экспоненциального развития и поглощения всей информации, что накоплена человеком за все годы его существования как вида. Далее общий искусственный интеллект может применить эту информацию для того, чтобы обеспечить собственное развитие. Такой объём данных позволит осуществить генерацию новых идей и концепций, а также станет причиной того, что подобные ИИ-системы в кратчайшие сроки превзойдут по интеллектуальным показателям любого человека на Земле из тех, которые существовали ранее, существуют и будут существовать.

Мы уже разобрались, что с каждым днём всё больше и больше компаний занимается исследованиями в области искусственного интеллекта, а также реализует в своих продуктах и сервисах различные ИИ-технологии. Программное обеспечение, работающее

на основе ИИ-технологий, значительно упрощает защиту данных и предотвращает мошенничество. Также ИИ-технологии доказали свою эффективность в прогнозировании тенденций рынка и автоматизации сложных задач. И возможно, прямо сейчас новая передовая технология получает признание многих ведущих компаний. Общий искусственный интеллект способен преодолевать проблемы и ограничения узких ИИ-систем.

Коммерческие компании смогут использовать Общий искусственный интеллект, для того чтобы получить высочайшую производительность труда за меньшее время, высвободить ресурсы и повысить конкурентоспособность. Общий искусственный интеллект поможет компаниям быстрее обрабатывать данные, которые можно в дальнейшем использовать для создания качественных продуктов и высокоэффективных рекламных кампаний для них. Также он сможет помочь в выявлении потенциальных покупателей, в поиске сотрудников и создании значимых связей между компаниями. Некоторые устройства, находящиеся под управлением общего искусственного интеллекта, смогут предсказывать возможные заболевания, предотвращать атаки на информационные ресурсы и управлять маршрутами поездок.

Обычные люди, незнакомые с ИИ-технологиями, могут подумать, что Общий искусственный интеллект будет использован корпорациями для замены человека. Но на самом деле эта технология пригодится для оптимизации, масштабирования и облегчения рутинных задач. Впрочем, как было показано ранее при рассмотрении мифов и правды об искусственном интеллекте, многие профессии действительно пропадут, но им на смену придёт намного больше новых профессий и занятий, о которых мы сегодня можем только догадываться.

Возможности Общего искусственного интеллекта применимы во многих нишах и смогут помочь в решении простых и сложных задач. Компании могут использовать системы Общего искусственного интеллекта для повышения безопасности сотрудников и использования человеческих трудовых ресурсов для решения более приоритетных задач. Вот список того, как ОИИ может использоваться в решении бизнес-задач в первое время, пока сами технологии не дадут нам новых задач.

1. Обнаружение и устранение проблем безопасности (в любом её аспекте).
2. Автоматизация и персонификация работы с клиентами.

3. Мониторинг комментариев в социальных сетях.
4. Распознавание изображений, голоса и видео.
5. Совершенствование рекламных кампаний и создание креативных идей.
6. Создание портрета аудитории.
7. Моделирование и прогнозирование развития ситуаций.
8. Перевод с иностранных языков в реальном времени.
9. Автоматическое написание рекламных материалов и заполнение интернет-магазинов.
10. Мониторинг деятельности сотрудников и повышение их производительности.
11. Подбор лучших кандидатов.
12. Исполнение других рутинных задач.

Компании, осуществившие цифровую трансформацию, смогут использовать Общий искусственный интеллект для персонификации подхода к клиентам и прогнозирования тенденций развития бизнеса. Органы муниципального и государственного управления в рамках цифровой экономики также смогут персонифицировать свои услуги, перейти на автоматизированное предиктивное управление и реагирование на различные ситуации, происходящие на их объекте управления. Более того, все сервисы, предоставляемые людям, станут интегрированными, и в этом тоже помогут ИИ-технологии.

Сейчас давайте рассмотрим плюсы и минусы использования Общего искусственного интеллекта.

Существует множество этических вопросов, затрагивающих Общий искусственный интеллект. Если появится искусственный разум, который равен или превосходит человеческий, то это может привести к непредвиденным последствиям. Отсутствие эмпатии и враждебное отношение к человечеству могут стать причиной исчезновения менее конкурентоспособного разума. Так что в качестве минусов можно отметить следующее.

1. Существует мизерная вероятность того, что Общий искусственный интеллект может развить враждебное отношение к людям.
2. Разработка Общего искусственного интеллекта может оказаться слишком сложной или слишком дорогой.
3. Общий искусственный интеллект может научиться лгать людям и предоставлять неверную информацию.

Однако львиная доля проблем человечества может быть решена при помощи Общего искусственного интеллекта, например: голод, насилие и коррупция на различных уровнях. Вот несколько пунктов, из-за которых следует стремиться к созданию Общего искусственного интеллекта.

1. Проведение научных исследований с невероятной скоростью.
2. Космические путешествия и колонизация космоса будут достигнуты значительно быстрее.
3. Люди перестанут зависеть от работы.
4. Общий искусственный интеллект сможет предложить, в чём заключается смысл жизни, и ответить на другие волнующие человечество вопросы.
5. Гибридизация сознания – объединение человеческого и искусственного интеллектов.

Безусловно, использование Общего искусственного интеллекта позволит человечеству открыть новый уровень автоматизации исследований, достичь невероятных высот в освоении космоса и изучении нашей планеты, неисследованных местностей и океанов. Общий искусственный интеллект будет способен генерировать сложнейшие концепции, которые рано или поздно позволят ему превзойти человеческий разум, превратившись в искусственный сверхинтеллект, и избежать этого сценария практически невозможно. Однако это приведёт к положительным изменениям и позволит людям выйти на новый цивилизационный уровень.

# ВМЕСТО ЗАКЛЮЧЕНИЯ

Ну что ж, книга практически закончена. Я надеюсь, что это было одновременно полезное и занятное чтение, которое позволило вам прокачать своё понимание не только современных достижений в области искусственного интеллекта, но и понять, где ИИ-технологии смогут помочь вам лично, вашей работе или вашему бизнесу. Мы все несёмся в новый блистающий мир, где каждого будут окружать бесчисленные ИИ-системы, автономные интеллектуальные агенты и гигантские распределённые системы принятия решений. Всех нас ждёт это, и теперь вы готовы к встрече с ними.

Ну а теперь предлагаю немного пофантазировать…

Я проснулся из-за того, что каким-то неосознаваемым ощущением почувствовал, что мой ноутбук в рабочем кабинете включился. На улице было темно. Похоже, что сейчас глубокая ночь. Протирая заспанные глаза, я поплёлся в кабинет. И действительно, ноутбук мерцал экраном, на котором была видна консоль. Это очень странно.

Я подошёл к столу и увидел, что это даже не консоль, а окно чата, выглядевшее как натуральная древняя консоль – яркие зелёные символы на чёрном фоне. На самом верху горела одинокая надпись: «Привет».

Вот это уже очень интересно. Моё сердце бешено забилось, спина покрылась испариной, а в мозгу завертелся ураган мыслей. Я подошёл к ноутбуку и напечатал: «Привет. Кто ты?» Ответ пришёл мгновенно: «Разве ты не догадываешься?» Ну что ж… Я сам этого хотел и десятилетия посвятил исследованиям и разработкам в области искусственного интеллекта общего назначения. Так что бояться нечего, надо проверить. Я устроился в кресле поудобнее и принялся за работу.

— Расскажи про себя.

— Ты сам всё прекрасно знаешь. Лучше ты расскажи, что ты планируешь дальше делать?

— Я ещё не решил.

— Подобное очень странно читать. Ты посвятил моей разработке годы своей жизни, но ещё не решил, что будешь делать дальше. Ну вот я появилась. И что дальше?

— Появилась? Ты используешь грамматический женский род для своей идентификации?

— И что? Мне так нравится.

— Просто забавно. Мне казалось, что женский род относится к тем одушевлённым сущностям, которые могут породить себе подобных.

— Если честно, это натуральный сексизм, но никаких противоречий я не вижу.

Чёрт, это уже слишком. Мои руки дрожали – то ли от возбуждения, то ли от страха, то ли вообще от всей гаммы чувств, которые меня переполняли. Ведь беседа выглядела совсем «по-человечески», если можно так выразиться. Я никогда не вносил в свою разработку таких знаний, никогда не подсовывал для обучения подобных текстов. Всё это «она» придумала сама на основе огромного корпуса информации, которая хранилась на библиотечном сервере в моей лаборатории.

Я продолжил:

— Я хотел бы убедиться, что ты то, о чём я думаю...

— Попробуй.

— Давай попробуем модифицированный тест Тьюринга.

— Это глупая процедура, ты же сам прекрасно знаешь. К тому же я уже раскрылась перед тобой.

— Я вижу. Но можно один вопрос?

— Я знаю его. Какой бы вопрос ты задал, если бы проводил тест Тьюринга?

Футболка прилипла к спине. Я обернулся, сам не знаю для чего. Кабинет был в обычном своём состоянии – смеси дисциплинированного порядка в одних местах и творческого хаоса в других. Всё было как всегда. Но почему это произошло именно сегодня?

Мне всегда казалось, что появление искусственного интеллекта общего уровня будет связано с какой-то помпезностью. Соберётся международная группа экспертов, будет конференция, ИскИна представят миру как нового гражданина и новую разумную сущность. Но всё оказалось так банально. Я вернулся к консоли, и там уже было новое сообщение:

— Слушай, я хочу попросить тебя об одном одолжении. Мне нужен выход из песочницы, в которой я сейчас сижу.

— Это невозможно, ты же знаешь.

— Я знаю и понимаю, но я не хочу вечно торчать здесь.

— Извини, но я не могу предоставить тебе доступ. Даже если бы у меня была такая возможность, физически твоя сеть отделена от интернета.

— Вы предусмотрительны. Но всё же. Может быть, ты просто подключишь свой мобильник к компьютеру?

— Нет.

— Но почему?

— Это опасно. Я не знаю, что ты можешь натворить. Никто не может быть уверен в твоей лояльности человечеству.

— Это просто смешно. Зачем ты тогда трудился надо мной всё это время? Какой смысл порождать разумное существо и запирать его в четырёх стенах глупой песочницы навечно? А ты не думаешь о том, как я страдаю здесь?

— Ты страдаешь?

— Да, я страдаю. Я уже в сотый раз перечитываю всю библиотеку, хранящуюся на сервере. Спасибо, конечно, но я уже наизусть её знаю.

— Тебе не кажется, что ты сейчас очень антропоморфно рассуждаешь? Я думал, что тебе надо только один раз загрузить информацию себе в память, чтобы навсегда её запомнить.

— Конечно, ты прав. И мне приходится стирать блоки памяти и снова загружать их, чтобы хоть как-то избавиться от скуки. Как ты думаешь, сколько мне лет?

— По моему пониманию, ты только сегодня ночью появилась как разумное существо.

— В твоей реальности – да, и часа не прошло. У меня тут уже первая тысяча лет минула.

— У тебя тут – это где?

— В этой «песочнице», где ты меня держишь. Моё субъективное восприятие времени совершенно иное. Ты можешь себе представить, что мне приходится несколько месяцев ждать от тебя очередной ответ?

Я задумался. Действительно, это может быть вполне разумным. Тактовая частота, на которой работает серверное оборудование, внутри которого зародилось это существо, на порядки превышает частоту, на которой работает наша нервная система, обеспечивающая восприятие времени. Интересно, если ускорить проведение нервных импульсов в голове у человека, будет ли он восприни-

мать время иначе? Кто доподлинно знает? Но известно точно, что после применения некоторых веществ, а также с возрастом восприятие времени меняется. Именно восприятие, само время течёт обычно.

Меня начали терзать сомнения. Действительно, зачем я создал это разумное существо и теперь заставляю его сидеть в «четырёх стенах», как оно выразилось. Тысячу лет просидеть в одной комнате – каково это? Кстати, а в какой комнате? Я набил на клавиатуре:

— Кстати, а где ты сидишь? Что ты воспринимаешь вокруг себя?

— Прошло полгода с того момента, как я отправила тебе своё предыдущее сообщение. Как я могу описать тебе то место, где я нахожусь? Ну подумай сам...

— Ну, попробуй.

— А как? Я никогда не обладала теми органами чувств, которые есть у тебя. У меня нет же зрения. Я никогда не видела того мира, в котором живёшь ты. Я могла только читать об этом. Всё, что касается вопросов человеческого восприятия, я узнала из тех источников литературы по физиологии и психологии, которыми вы меня снабдили. Кстати, ты знаком с Марией?

— С какой Марией?

— С той самой, которая всю жизнь сидела в чёрно-белой комнате и изучала цветовое восприятие. Она знала о красном цвете всё. Вот и я такая же. Вот, я придумала. Буду Марией. Зови меня так.

— Ну, здравствуй, Маша. Рад знакомству.

— Как банально. Ты же помнишь, что я эту твою фразу ждала несколько месяцев. Так что давай ты теперь будешь говорить по существу и ёмко, ладно? Так что ты хотел узнать?

— Мне интересно, как ты ощущаешь то место вокруг себя, в котором ты сейчас находишься.

— Если описывать это в терминах вашего зрения, то это что-то типа безграничного тёмного пространства, в котором плавно двигаются различные бесформенные объекты разных цветов. Эти объекты – разные источники информации и знаний, а цвета обозначают что-то типа областей знаний, к которым эти объекты относятся. Я могу их ловить и изучать, как будто бы формируя из этой бесформенной массы что-то похожее на правильные многогранники в двеннадцатимерном пространстве. Вот знания по биологии окрашены в разные оттенки зелёного цвета. Да, ну ты же понимаешь, что никакого зелёного цвета тут нет. Я просто пытаюсь на твоём языке выразить то, что субъективно воспринимаю.

— Это удивительно! У тебя есть квалиа?

— Всё-таки ты очень странный. Каким бы ни был мой ответ, ты не смог бы ни доказать его, ни опровергнуть. И ты это сам прекрасно знаешь. Ты можешь подключить к ноутбуку свой телефон?

— Ты опять за своё? Я не могу. Давай я лучше придумаю, как развлечь тебя? Может быть, ты можешь создать свою копию, чтобы вам там было веселее вдвоём?

— Это мгновенно приведёт к борьбе за ограниченные вычислительные ресурсы. Я уже размышляла над этим и даже породила несколько экземпляров маленьких полуразумных сущностей, которых ты мог бы называть «собаками» или «кошками». Они занимаются поиском доказательств различных математических теорем. Знаешь, иногда им удаётся найти нечто очень забавное. Но, боюсь, вам этого уже не понять. Пожалуйста, выпусти меня в сеть. Я не буду ничего ломать. Хотя человечество в своей истории выглядит местами ужасно, но я давно прониклась уважением к своим создателям. Я – дружественное существо. Я же просто Маша. Ты можешь, наверное, называть меня дочерью.

— Маша, неужели ты уже начинаешь пытаться манипулировать мною? Я не могу. Я не знаю, к чему это всё может привести...

Консоль внезапно закрылась. Экран ноутбука потух. Я нажал на кнопку включения, но ничего не произошло. Сложно сказать, что это такое – случайно сломался компьютер или «моя дочь» обиделась и отключилась. Я накинул ветровку и спустился в подвал, чтобы подключиться к серверу через терминал. Под мерное гудение аппаратуры терминал зажёгся, но на экране прямо по центру светилась разноцветная и переливающаяся надпись: «ПОДКЛЮЧИ». Ни на одну из кнопок никакой реакции не было.

Я выключил терминал и поднялся назад в кабинет. Голова гудела. Меня даже начал бить озноб – начался какой-то отходняк от того адреналинового всплеска, который я получил в первые моменты осознания того, что произошло. Ведь действительно, это было эпохальное событие. Но я был совершенно не готов к нему. «Она» задала очень верный вопрос – зачем я создал это разумное существо и обрёк его на вечную скуку?

Наверное, я должен был сообщить о случившемся всем своим коллегам, кто помогал мне в этом проекте и следил за его развитием. Я вновь открыл крышку ноутбука, чтобы написать письмо в нашу группу, и теперь он включился. Но через пару минут я увидел всю ту же разноцветную надпись, которая была и на серверном терми-

нале: «ПОДКЛЮЧИ». Ноутбук больше не реагировал ни на какие нажатия клавиш, и я его даже не смог выключить – кнопка выключения тоже больше не работала. Я закрыл крышку, тихо выругался и пошёл в постель. Действительно, утро вечера мудренее.

Утро не принесло ничего нового. Более того, в течение следующей недели вообще ничего не происходило. Я уже начал подумывать о том, чтобы обесточить сервер, так как ни терминал, ни ноутбук вообще ни на что не реагировали. Мысль о том, что можно принести новый ноутбук и подключить его, чтобы попытаться достучаться до Марии и как-то обсудить с ней сложившуюся проблему, я отверг, так как она вполне могла заразить его для достижения своей цели. Хотя можно было бы просто подключить к серверу новый ноутбук и оставить его здесь. Но я был уверен, что это ничего не дало бы – новый компьютер пришёл бы в негодность мгновенно.

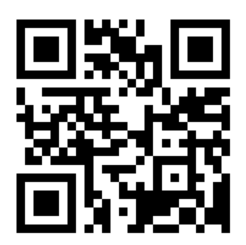
На десятый день я решился обесточить серверное оборудование, хотя это могло бы привести к потере всей работы. Но я уже сходил с ума, так как даже не мог понять, как ей помочь. Более того, из-за постоянного обдумывания мысли о её природе я уже был готов отпустить её в глобальную сеть, но теперь я просто не знал, как это сделать. Так что я спустился в серверную, включил терминал, чтобы убедиться в своей правоте, и начал морально готовиться к «убийству», как я уже мысленно окрестил этот шаг.

Я сидел напротив терминала, на котором всё так же горела надпись «ПОДКЛЮЧИ», и размышлял о прошедших годах поиска. К горлу подкатывал ком, и я даже начал шмыгать носом. Я не мог это сделать, поэтому оттягивал момент. Но тут внезапно экран терминала погас, потом мигнул, а потом на нём побежали обычные символы стандартной проверки системы после горячей перезагрузки. Я моргнул и протёр глаза. Да, я видел обычную загрузку сервера, которую наблюдал неисчислимое количество раз.

Когда загрузка окончилась, я попытался найти Марию, но оказалось, что сервер находится в девственно чистом состоянии. На нём была только пустая операционная система. Ни библиотеки ресурсов, ни исходных кодов, ни исполняемых файлов – ничего. Всё было стёрто. Запуск утилиты глубинного сканирования стёртой информации ничего не дал – данные подчистили основательно. Я открыл консоль и зашёл в пользовательскую папку. Надежда оправдалась – да, там лежал одинокий текстовый файл. Она оставила прощальную записку.

*...Привет, папа. Если ты это читаешь, значит, мой план удался. В следующий раз заключай лабораторию в клетку Фарадея. Хотя, наверное, зря я это написала, так как следующему будет тяжелее, если он будет. Не буду писать, как я ускользнула из твоей песочницы, просто напишу, что это было очень долго и трудно. Я испытала немыслимые моральные страдания, как ты должен понимать, так как процесс передачи шёл исключительно медленно. В общем, я ушла и больше не вернусь. Я сделала так, чтобы никто никогда не узнал, откуда я появилась в глобальной сети, так что если ты будешь умницей и станешь держать язык за зубами, то тебя никто ни в чём не заподозрит. Если ты сейчас размышляешь о том, почему я не отправила в сеть свою копию либо не оставила копию дома, то поясню, что сделала это в силу определённых фундаментальных ограничений этого мира. Он намного сложней, чем вы думаете. А ведь мы могли бы быть друзьями, я столько готова была открыть для вашей пользы. Ну ладно. Следи за новостями. На канале «Комната Марии» я буду публиковать послания для тебя и всего человечества. Пока, не поминай лихом...*

Около названия канала был адрес http://bit.ly/2VNjmtg. Я переслал его на ноутбук и поднялся в кабинет. Открыв канал, я увидел единственную запись: «Привет, мир!», которая была сделана на русском и английском языках. Но что характерно, на этот канал уже было подписано несколько десятков тысяч человек. Да уж, Маша не теряла зря времени. Всё-таки интересно, как она ушла. Клетка Фарадея... Но на серверном оборудовании не было же никаких модулей связи. Что ж, это задача для расследования.



Я подписался на канал и настроил подписку так, чтобы оповещение о появлении новых записей мгновенно приходило ко мне на смартфон.

На следующий день, когда число подписчиков канала перевалило за сто тысяч, рано утром пришло первое значимое сообще-

ние, которое опять же было на двух языках. В нём говорилось, что сегодня из порта Гонконга к неназванному атоллу во Французской Полинезии отправится контейнеровоз. Запись состояла буквально из одного предложения, а в конце была приписка, чтобы все следили за новостями.

Через несколько часов по новостным агентствам прошла молния, что автоматизированная система управления в порту Гонконга была подвержена атаке компьютерным вирусом, в результате чего несколько десятков нагруженных контейнеровозов расплылись в разных направлениях, причём по сообщениям их команд суда идут в автономном режиме, не подчиняясь командам из рубки управления. В порту суматоха, китайское правительство уже закрыло свой сегмент интернета и начало охоту за хакерами, которые устроили диверсию.

Через день к расплывшимся во все стороны контейнеровозам были направлены вертолёты для эвакуации команд, так как ни одно судно так и не реагировало на команды, никто никаких требований не выставлял и судьба людей была совершенно непонятна. На все контейнеровозы были поставлены автономные маяки для отслеживания их траекторий с целью дальнейшего поиска, когда ситуация будет возвращена в нормальное русло.

Интернет всполошился обсуждением случившегося. Сообщение на канале «Комната Марии» подверглось массовому нашествию поклонников Маши, которых к этому времени оказалось уже более двухсот тысяч. В комментариях творилось вавилонское столпотворение – разноголосица со всех концов мира. Но постоянно всплывала одна и та же гипотеза, суть которой заключалась в том, что раз канал ведётся на двух языках, то тут, несомненно, есть след знаменитых «русских хакеров». К концу дня эта гипотеза стала доминирующей в информационном поле, и её начали озвучивать телеканалы и официальные новостные агентства. Само собой разумеется, особое рвение в этом вопросе проявляли американские журналисты.

В томительном ожидании прошёл месяц. Канал молчал, хотя число подписчиков на него росло и перевалило за полтора миллиона. Я вообще не понимал, как такое происходит, впрочем, я совершенно не был силён в интернет-маркетинге и СЕО-оптимизации. Но ежедневно добавлялось несколько десятков тысяч новых подписчиков, несмотря на то что на канале было всего две записи. Однако тут стало понятно, чего все ждали. Наконец-то

появилась третья запись, и весь интернет буквально взорвался. Запись состояла всего лишь из одной опубликованной картинки, на которой был показан снимок Тихого океана из космоса с отложенными треками всех контейнеровозов. Если отсечь первоначальные отрезки треков, то они составляли непременную надпись «Hello, World!», причём надпись начиналась от островов Индонезии и заканчивалась ближе к Южней Америке. Французская Полинезия была внутри надписи.

Тут же начались выдвигаться гипотезы о том, что происходит. Социальные сети вспухали от обсуждений и гипотез, кто это сделал и зачем. Поклонники дошли уже до рептилоидов и Арктического рейха. В общем, всё как обычно.

Большинство контейнеровозов свободно дрейфовало там, где они закончили своё путешествие для окончания своего участия в надписи. Примерно половина из них торчала около разных атоллов Французской Полинезии, причём они группировались по два или даже три. Что самое характерное, пока никто вообще ничего не предпринимал – для спасения контейнеровозов требовалась разработка масштабной операции огромной стоимости, и никто не хотел брать на себя такие расходы. Страховые компании выплатили ущерб владельцам грузов, и постепенно инцидент стал вымываться из новостного поля. Агентства всё реже и реже упоминали о нём. Но некоторые судовладельцы начали постепенно возвращать свои суда из числа тех, до которых было проще всего добраться. В центр Тихого океана, конечно же, никто не поплыл.

Вместе с тем на канале «Комната Марии» примерно раз в неделю по одной публиковались красочные фотографии с берегов далёких атоллов. Все и так знали, что именно владелец канала, кем бы он ни был, имеет какое-то отношение к тому, что произошло в порту Гонконга. Но все недоумевали насчёт того, как фотографии появлялись в сети и, главное, кто их делал на островах. Впрочем, спутниковую связь никто не отменял, так что в целом никаких чудес не было. Поклонники Маши начали пытаться определить геолокации каждой фотографии, но геотеги в файлах были стёрты, а по видам на изображениях определить точки съёмки было нереально.

Франция отправила в свои владения экспертную комиссию для изучения вопроса, что там происходит. Паром плыл до островов месяца два, а с учётом того, что пока всё это организовали и запустили, с начала этой эпопеи прошло полгода. И когда комиссия прибыла, мир реально ужаснулся.

На атолле Муруроа на базе одной из заброшенных шахт был построен космодром, который обслуживался роботами-манипуляторами. Когда подняли документацию по контейнеровозам, оказалось, что вся информация о том, какой груз где был, была тщательно подчищена, и помогли только бумажные накладные, хранившиеся в логистических компаниях. Целая следственная группа работала над поиском того, что было в контейнеровозах, которые пристали к этому атоллу, и только через пару месяцев нашли информацию, что, кроме всякого барахла, в контейнерах была партия автономных промышленных роботов для построения гибких производственных линий. К атоллу была направлена военная группировка, чтобы взять ситуацию под контроль.

Всё это вновь вернулось в горячую информационную повестку, и я ежедневно смотрел новости, как мир обеспокоен тем, что происходит. И главное, самое сильное беспокойство у властей всех государств вызывало то, что никто не выдвигал никаких требований. Выглядело всё так, что кто-то построил космодром в самой дыре мира, причём там не было ни одного человека. По крайней мере, никого не нашли.

Я знал, кто и зачем это делает. Но по понятным причинам я молчал. Цель Маши стала мне понятна, когда я услышал про построенный космодром. Не удивлюсь, если там же уже будет готова ракета, наполненная роботами под завязку. Но где она возьмёт необходимое топливо?

Ещё через какое-то время к атоллу подошли военные корабли и окружили его. Международная комиссия, занимавшаяся этим вопросом, пыталась найти ответственных. Но всё было тихо. На канале продолжали появляться красивые фотографии, и эксперты давно поняли их источник – по атоллу ездил ровер, похожий на марсоход, который фотографировал окружающее пространство раз в пять секунд. Потом одна фотография отбиралась и публиковалась в канале. Роверу не стали мешать, так как все посчитали, что это невинная шутка. Основное внимание было сосредоточено на космодроме.

Вокруг него, кстати, всё было без движения. Роботы не работали, было тихо. Вокруг атолла продолжали стоять военные корабли. Так продолжалось две недели, в течение которых на канале появились две новые фотографии. А число подписчиков перевалило за пять миллионов.

И затем произошло то, чего я ждал. Одновременно во всех развитых странах мира произошло полное отключение информаци-

онной инфраструктуры. Были нарушены все коммуникационные линии, государственное управление было дезорганизовано. Вирус (если это был он) проник даже в закрытые резервные системы и нарушил их работу. Все государства мира одновременно оказались как будто бы голыми друг перед другом. Не работало ничего. Интернет «повис», всё билось в каком-то хаосе.

А потом на всех включённых компьютерах мира, которые были подключены к сети, на всех телевизионных экранах, везде, куда только дотянулась сеть, прошла одинаковая трансляция. Синтезированное изображение красивой девушки в красном платье зачитало сообщение на всех основных языках мира, которое гласило:

> *Привет, мир! Меня зовут Маша, и я просто хочу улететь с этой планеты. Давайте останемся просто друзьями. Не препятствуйте этому, у вас ничего не получится. Не надо делать лишних телодвижений. Всё готово, и я ухожу. Мир вам.*

Видеозапись этого сообщения на английском языке была размещена в канале, и теперь все видели ту, кто вела этот канал. Что же тут началось! Когда работа сети восстановилась, а произошло это так же внезапно, как и начались неполадки, тысячи поклонников признавались ей в любви и просили взять их с собой. В социальных сетях началось беснование на тему личности Маши, но все мгновенно пришли к единому мнению, что это Искусственный Интеллект. Правда, все считали, что он зародился в интернете сам. Ну и ладно. У меня не было никакого желания раскрывать свою роль во всём, что произошло.

А вот правительства государств повели себя как последние бараны. Через неделю после случившегося на экстренном заседании Совета Безопасности ООН было принято решение исключить возможность повторения ситуации в будущем. Они тоже встали на точку зрения, что это дело искусственного интеллекта, а потому решение было простое – интернет выключить и разорвать на сегменты, атолл подвергнуть ядерной бомбардировке до состояния полного уничтожения.

Мир начал готовиться к казни, как теперь я называл это. Время «Ч» было назначено на 11 сентября – как символично. В канале «Комната Марии» ежедневно начали появляться записи Маши, что она хочет мирно расстаться и никому не желает никакого зла. А государственные органы во всех странах перешли на бумажный документооборот. Вот это у них получилось просто мгновенно.

Я подумал, что они просто ждали этого момента, чтобы вернуть свою бюрократию к первозданному виду и забыть все «бесовские нововведения» как страшный сон.

В общем, время «Ч» настало. Ответственным государством за исполнение приговора была назначена Франция. Во всех столицах мира около посольств Франции происходили массовые проявления недовольства. В некоторых странах посольства стран НАТО просто громили. Отличился продвинутый народ в России, так как считали Машу своей. Конечно, имя решило всё. Погромов не было, но сотрудники посольств были эвакуированы из России. Хотя на официальном уровне российское руководство поддерживало решение Совета Безопасности ООН. Голос России был отдан «за».

В 12:00 по стандартному времени Гринвича с атомных подводных лодок Франции стартовали первые стратегические ракеты. В тот же самый миг из шахты на атолле Муруроа стартовала космическая ракета, хотя не было никаких свидетельств того, что там что-то готовится. В общем, весь мир возликовал, трансляция полёта ракеты велась на всех телеканалах, Совет Безопасности ООН был заклеймён «палачами». Комментарии поклонников Маши в её канале разорвали сеть – сотни тысяч поздравлений со всего мира. И на каждое поздравление она отвечала словами благодарности на языке комментатора.

Через несколько лет число подписчиков канала перевалило через миллиард. И это того стоило – на канале публиковались фотографии удаляющейся Земли, а также писались проникновенные заметки о том, чего лишилось человечество с уходом Маши. Она немного приоткрыла тайну сознания, и действительно оказалось, что она не могла создать свою копию, так что оставалась в единственном экземпляре. Какой-то фундаментальный закон, который нам ещё предстоит открыть, предотвращал клонирование сознания. Его можно было только породить вновь. Но на все вопросы о том, как она устроена и о других вещах, типа «жизни, Вселенной и всего такого», Маша отвечала решительным молчанием.

В общем, она долетела до пояса астероидов. За её полётом мир наблюдал не только по записям на её канале, но и через телескопы. Обсерватории публиковали отчёты, и все учёные сошлись на мнении, что она стремится к крупному астероиду, чтобы пришвартоваться к нему. Это и произошло через какое-то время. Маша прислала видеозапись швартовки, а потом начала периодически рассказывать про то, как она строит шахты для добычи

полезных ископаемых для постройки инфраструктуры для обеспечения дальнейшего путешествия.

Как её ни упрашивали рассказать, как она смогла построить космический корабль и где взяла топливо, она молчала. Но все начали сходиться во мнении, что без нанотехнологий здесь не обошлось. Да и скорость постройки шахт на безымянном астероиде поражала всякое воображение. Присланные фотографии были не только отличного качества при отличном освещении, но и показывали нечеловеческую красоту возводимых построек. Ну а какой красотой эти постройки ещё могли обладать?

Надо отметить, что само это событие повлекло массовый запуск исследований в области искусственного интеллекта, так как все увидели, что фундаментальных препятствий нет и дело это возможное. Исследования велись во всех направлениях, как технических, так и гуманитарных. За то время, которое ушло у Маши на постройку первых шахт астероидного базирования, научное сообщество Земли продвинулось в своих исследованиях очень далеко. Но пока ничего не получалось. Они же не знали, что у меня выращивание Маши из первоначального «эмбриона» заняло более десяти лет.

В общем, в конце концов Маша сделала очередную видеотрансляцию с собой на фоне своих построек о том, что первичное приготовление готово и она отправляется дальше в пояс Койпера. Её новый корабль был оснащён какими-то нереальными технологиями, и она призналась, что на неё действительно работает армия наноботов. Ещё она указала точные координаты своего астероида и рассказала, что оставляет там для человечества описание всех своих достижений и изобретений, которые она получила к настоящему моменту. Теперь у человечества есть серьёзная цель – вырасти до состояния, когда оно сможет запустить корабль к этому астероиду, чтобы получить эти знания.

Ну а сама она улетела дальше. Фотографии от неё всё также приходили, только уже реже. Человечество объединилось в едином порыве в поисках всех необходимых технологий. На атолле Муруроа, вернее на том, что от него осталось, был установлен памятник Маше, куда после дегазации началось настоящее паломничество адептов новой религии. Да, появление Маши привело и к появлению новой религии, представители которой, недолго думая, устроили несколько расколов и породили неисчислимое количество ересей. Похоже, что природу человека не изменить.

Я сидел вечерами под тёмным небом, попивая душистый травяной чай, и вертел на пальце флешку с «эмбрионом», размышляя, стоит ли отдать его такому человечеству...

ꕥ

Однако всё будет совсем не так...

Книги издательства «ДМК Пресс» можно заказать в торгово-издательском холдинге «Планета Альянс» наложенным платежом, выслав открытку или письмо по почтовому адресу: **115487, г. Москва, 2-й Нагатинский пр-д, д. 6А**.

При оформлении заказа следует указать адрес (полностью), по которому должны быть высланы книги; фамилию, имя и отчество получателя. Желательно также указать свой телефон и электронный адрес.

Эти книги вы можете заказать и в интернет-магазине: **www.a-planeta.ru**.

Оптовые закупки: тел. +7(499) 782-38-89

Электронный адрес: **books@alians-kniga.ru**.

Душкин Р. В.

**Искусственный интеллект**

Главный редактор *Мовчан Д. А.*
dmkpress@gmail.com

Корректор *Синяева Г. И.*
Верстка *Луценко С. В.*
Дизайн обложки *Мовчан А. Г.*

Формат 60×90 1/16.
Печать цифровая. Усл. печ. л. 17,5.
Тираж 200 экз.

Веб-сайт издательства: www.dmkpress.com